DMD Fast Forward. Published on June 8, 2006 as doi:10.1124/dmd.105.008854

DMD MS#8854

GENETIC VARIABILITY, HAPLOTYPE STRUCTURES, AND ETHNICAL DIVERSITY OF HEPATIC TRANSPORTERS MDR3 *(ABCB4)* AND BSEP *(ABCB11)*

Thomas Lang, Michael Haberl, Diana Jung, Anja Drescher, Robert Schlagenhaufer, Andrea Keil, Esther Mornhinweg, Bruno Stieger, Gerd A. Kullak-Ublick and Reinhold Kerb

EPIDAUROS Biotechnologie AG, Am Neuland 1, 82347 Bernried, Germany (T.L., M.H., A.D., R.S., A.K., E.M., R.K.), Department of Pharmacology, Johannes Gutenberg University, 55101 Mainz (T.L.), Germany, Division of Clinical Pharmacology and Toxicology (D.J., B.S., G.A.K.-U.) and Division of Gastroenterology and Hepatology (G.A.K.-U.), Department of Internal Medicine, University Hospital Zurich, Zurich, Switzerland

Running title: **SNPs, HAPLOTYPES, AND ETHNICAL DIVERSITY OF *ABCB4* AND *ABCB11***

Corresponding author:

Reinhold Kerb, MD
Medical Science
AstraZeneca R&D
Pepparedsleden 1
SE-43183 Mölndal
Sweden

Phone: +46-31-7064590
Fax: +46-31-7063844

e-mail: reinhold.kerb@astrazeneca.com

Number of text pages:	26
Number of tables:	5
Number of figures:	6
Number of references:	34
Number of words in the Abstract:	243
Number of words in the Indroduction:	512
Number of words in the Discussion:	1651

Abbreviations: MDR, multidrug resistance; BSEP, bile salt export pump; DIC, drug-induced cholestasis; PFIC, progressive familial intrahepatic cholestasis; ICP, intrahepatic cholestasis of pregnancy; PBC, primary biliary cirrhosis; PSC, primary sclecrosing cholangitis; SNP, single nucleotide polymorphism; LD, linkage disequilibrium; PCR, polymerase chain reaction; ABC, ATP-binding cassette; AA, AfricanAmerican; CA, Caucasian; JA, Japanese, KO, Korean.

Abstract:

Biliary excretion of bile salts and other bile constituents from hepatocytes is mediated by the apical (canalicular) transporters P-glycoprotein 3 (MDR3, *ABCB4*) and the bile salt export pump (BSEP, *ABCB11*). Mutations in the *ABCB4* and *ABCB11* contribute to cholestatic diseases (e.g. PFIC2, PFIC3 and ICP) and our objective was to establish genetic variability and haplotype structures of *ABCB4* and *ABCB11* in healthy population of different ethnic background. All coding exons, 5 of 6 non-coding exons, 50-300 bp of the flanking intronic regions and 2.5-2.8 kb of the promoter regions of *ABCB4* and *ABCB11*, were sequenced in 159 and 196 DNA samples of Caucasian, African American, Japanese and Korean origin. Totally, 76 and 86 polymorphisms were identified in *ABCB4* and *ABCB11*, respectively, among them 14 and 28 exonic polymorphisms, 8 and 10 protein-altering variants, of which 4 were predicted to have functional consequences. Both genes showed substantial ethnic differences with respect to allele number, frequency of common and population-specific sites, and in patterns of linkage disequilibrium. Population genetic analysis suggested some selective pressure against changes in the protein supporting the important endogenous role of these transporters. Haplotype variability was greater in *ABCB11* than in *ABCB4*. An *ABCB11* promoter haplotype was associated with significant decrease of activity compared to wild type. Our results contribute to a better understanding of the molecular basis and of ethnic differences in drug response and provide a valuable tool for in future research on the hereditary of cholestatic liver injury.

Introduction:

ATP-binding cassette (ABC-) transporters mediate energy-dependent transport of exogenous and endogenous organic compounds across membranes with distinct substrate specifity, tissue distribution and intracellular localization (Borst and Elferink, 2002). The clinical importance of ABC transport proteins and the consequences of genetic polymorphisms is being increasingly appreciated. For instance, the P-glycoprotein or multidrug resistance protein 1 (MDR1), which was initially shown to confer resistance to cancer chemotherapy, has a considerable impact upon disposition and therapeutic response of many drugs. Single nucleotide polymorphisms (SNPs) in the *MDR1 (ABCB1)* gene have been reported to modulate expression levels, protein activity and bioavailability of substrate drugs (Kerb et al., 2006). Like MDR1, MDR3 and the bile salt export pump (BSEP) are members of the *ABCB* gene family. They are localized in the canalicular membrane of hepatocytes where they form the secretory biliary unit of the liver. BSEP is the predominant hepatocellular efflux system for the excretion of conjugated bile salts, while MDR3 acts as a flippase and translocates phosphatidylcholine across the canalicular membrane (Borst and Elferink, 2002; Kullak-Ublick et al, 2004). A lack of this major phospholipid in bile leads to formation of toxic monomeric bile salts in the bile ducts.

MDR3 and BSEP are potentially important targets for drug-induced liver injury. MDR3 P-gp is able to transport a number of MDR1 P-gp substrates (e.g. digoxin, paclitaxel and vinblastin), although the contribution of MDR3 is probably clinically less important. Verapamil, cyclosporine and vinblastine are able to inhibit MDR3, explaining why these drugs could adversely affect canalicular phosphatidylcholine secretion (Smith et al., 2000). Inhibition of BSEP by several compounds including estrogen, rifampicin, cyclosporine, troglitazone, and bosentan has been implicated in drug-induced cholestasis (DIC) (Stieger et al., 2000; Funk et al., 2001; Borst and Elferink, 2002; Kostrubsky et al., 2003; Kullak-Ublick et al., 2004).

Moreover, defects in the genes coding for MDR3 and BSEP, *ABCB11* and *ABCB4*, can cause hereditary disorders of the liver such as progressive familial intrahepatic cholestasis subtypes 2 (PFIC2, Byler syndrome) and 3 (PFIC3) or intrahepatic cholestasis of pregnancy (ICP) as reviewed recently (Pauli-Magnus et al., 2003). Currently, little is known about common *ABCB4* and *ABCB11* polymorphisms and their frequencies in healthy subjects of different ethnic populations. Genetic variants in *ABCB4* and *ABCB11* have been identified mostly in small samples of patients suffering from PFIC2 and PFIC3 syndromes, intrahepatic cholestasis of pregnancy and healthy control samples (Pauli-Magnus et al., 2003; Pauli-Magnus et al., 2004a;b). Saito et al., (2002) investigated 48 Japanese individuals and identified 50 and 98 genetic variations in *ABCB4* and *ABCB11*, respectively. Among these were 8 exonic polymorphisms, including only one protein-altering mutation, which was found in *ABCB11*. To the best of our knowledge, no systematic screens for polymorphisms have been carried out for *ABCB4* in Koreans and *ABCB11* in African Americans.

We, therefore, screened different ethnic populations for genetic variations in *ABCB4* and *ABCB11* and present the haplotype structure, a pair wise linkage disequilibrium mapping of polymorphisms and report population-genetic parameters of *ABCB4* and *ABCB11*. Furthermore, we investigated promoter activity of reporter gene constructs comprising polymorphisms in the 5' flanking regions of *ABCB4* and *ABCB11*.

Material and Methods:

DNA samples

Blood samples for DNA extraction were withdrawn from 149 Caucasian, 47 African American, 48 Japanese and 48 Korean healthy volunteers. 64 and 101 Caucasians, 47 and 48 Japanese, as well as 48 Koreans and 47 African Americans were sequenced for *ABCB4* and *ABCB11*, respectively. All subjects were healthy based on medical history, physical examination and routine laboratory tests. Anonymized blood samples were obtained for the purpose of pharmacogenetic testing after approval from the responsible local ethics committees and written informed consent. Caucasians were of German origin and selected in Berlin for participation in phase-1 clinical trials. African American DNA samples were purchased from Genomics Collaboratives Inc. (Cambridge, Massachusetts, USA). Japanese samples were purchased from Shin Nippon Biomedical Laboratories Ltd. (Kagoshima, Japan). Korean samples were purchased from the Department of Clinical Pharmacology (Pusan Paik Hospital, Inje University College of Medicine, Korea). The study was performed in accordance with ethical principles that have their origin in the Declaration of Helsinki.

DNA sequencing

For *ABCB4*, approximately 17 kb were screened including non-coding exons -3, -2, -1, 1 and coding exons 2-28 including 50 to 300 bp of adjacent intronic sequences, as well as 2610 bp of the promoter region upstream from the translation start in exon 2. In addition, 2841 bp of the 5' region upstream of the first untranslated exon -4 were screened. Approximately 13 kb of the *ABCB11* gene were screened including non-coding exon 1 and 2498 bp of the upstream promoter region and coding exons 2-28 and 50 to 300 bp of intronic sequence around these exons.

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

Genomic and cDNA sequences for primer design were derived from GenBank Accession Numbers AC005068.2 for promoter region, noncoding exons -3, -2, -1 and 1, coding exons 2 and 3, AC006154.1 for exons 4-12, AC0005045.2 for exons 13-28 and NM_000443.2 (cDNA) for *ABCB4*, and AC008177.3 for promoter and exon 1-21, AC069165.2 for exon 22-28 and NM_003742.2 (cDNA) for *ABCB11*. Exon-intron boundaries of the *ABCB4* and *ABCB11* gene were defined by comparing genomic sequences with the cDNA sequences. PCR reactions for generating *ABCB4* fragments were generally performed in a reaction volume of 50 µl with 20-30 ng of genomic DNA, 10X PCR buffer (Qiagen, Hilden, Germany), 200 µM dNTPs, 20 pmol of each primer and 1 unit Taq polymerase (Qiagen, Hilden, Germany). For amplification of exon 1, 4 µl DMSO (8%) were added, 1 µl 5X Q-solution (Qiagen, Hilden, Germany) were added for exons 2 and 3 and 2 µl of 25 mM $MgCl_2$ in addition to $MgCl_2$ in PCR buffer were added for exons 22 and 27. PCR reactions for generating *ABCB11* fragments were performed in a reaction volume of 50 µl with 20-30 ng of genomic DNA, 10X PCR buffer (Qiagen, Hilden, Germany), 200 µM dNTPs, 10 pmol (for promoter fragments Prom9, Prom8, Prom7, Prom6, Prom5, Prom4, Prom2 and exons 5, 8, 10, 12, 13, 14, 16, 17, 20, 21, 22, 23, 24, 26) and 20 pmol (for promoter fragment Prom3 and exons 1, 2, 3, 4, 6, 7, 9, 11, 15, 18, 19, 25, 27, 28) of each primer, 5 µl of 10 mM $MgCl_2$ in addition to the 15 mM $MgCl_2$ in PCR buffer and 1 unit Taq polymerase (Qiagen, Hilden, Germany) or 1 unit HotStarTaq polymerase (Qiagen, Hilden, Germany) for promoter fragments Prom9 and exons 12 and 22. PCR fragments were generated in a GeneAmp PCR System 9700 (ABI, Weiterstadt, Germany) with an initial denaturation step of 2 min at 94 °C, followed by 34 cycles of denaturation at 94 °C for 45 sec, annealing for 45 sec at 62 °C and extending for 1 min at 72 °C. The initial denaturation temperature for *ABCB4* fragments Prom2 and exons 1, 2, 3 was 96 °C for 3 min followed by 34 cycles of denaturation 96 °C for 45 sec, annealing for 45 sec at 62 °C and extending for 1 min at 72 °C. The initial denaturation temperature for *ABCB4* fragments

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

Prom9 and exons 12, 22 was 95 °C for 15 min followed by 34 cycles of denaturation 94 °C for 45 sec, annealing for 45 sec at 62 °C and extending for 1 min at 72 °C. Oligonucleotide primer sequences are given in Tables 1a and 1b. Subsequently purified amplicons were directly sequenced for genetic polymorphisms on PE ABI 3700 DNA Analysers by using BigDye Terminator cycle sequencing reactions (ABI, Weiterstadt, Germany). Sequences were analysed and polymorphisms identified using the PHRED/PHRAP/CONSED/POLYPHRED software package (University of Washington, Seattle, WA, USA). The sequences were inspected for deviations from *ABCB4* (NM_000443.2) and *ABCB11* (NM_003742.2) sequences, which were defined as reference. Singletons were confirmed by generating a second independent PCR fragment and direct sequencing from both ends.

Statistics, population genetic, structural and haplotype analysis

Computation of nucleotide diversity (π), neutral parameter (θ) and Tajima´s *D* (Tajima, 1989) was carried out using Arlequin 2.0 (Schneider et al. 2000). Functional tolerability of amino acid exchanges was calculated using SIFT (blocks.fhcrc.org/sift/SIFT.html), PolyPhen (www.bork.embl-heidelberg.de/PolyPhen), the Blosum62 amino acid substitution matrix (www.embl-heidelberg.de/~seqanal/courses/predoc97/blosum62.cmp) and Grantham values (Grantham, 1974). Amino acid residues were classified in being evolutionarily conserved or unconserved based on protein sequence alignments with mammalian orthologs by using SIFT and ClustalW (www.ebi.ac.uk/clustalw/). Three mammalian protein sequences (the human sequence and at least two from mouse, rat, and / or rabbit) were used for the alignments. An amino acid residue was classified as evolutionarily conserved (EC) if it was present in all members of a set of mammalian orthologs; all others were classified as evolutionarily unconserved (EU). Transmembrane domain (TMD) and loop regions were assigned based on topology data from SwissProt database (www.ebi.ac.uk/swissprot).

Allele frequencies for the *ABCB4* and *ABCB11* polymorphisms were compared pair-wise between population groups and analyzed for deviation from Hardy Weinberg equilibrium using Fisher's exact test. Linkage disequilibrium (LD) between pairs of polymorphisms was quantified by D'and r^2 statistics.

Rare mutations, which were only found once in the entire sample set, were excluded from haplotype analysis. Haplotypes and their frequencies were statistically calculated using PHASE 1.0 (Stephens et al., 2001) by running ten iterations with different seeds and default parameters. A haplotype was regarded as certain if it was inferred identical in at least 7 runs. PHASE can cope with missing genotype data (*ABCB4* 7.8%, *ABCB11* 7.8%). LD analyses were performed with the inferred haplotypes using DnaSP4.0 (Rozas et al., 2003). The HAPLOVIEW software package using the confidence intervals algorithm was employed to determine LD block structures (Barrett et al., 2005).

Promoter activities were statistically evaluated using SPSS 10.0.8 for Macintosh (SPSS Inc., Chicago, USA). The General Linear Model (GLM) Univariate procedure was applied to compare luciferase activities of the mutant promoter constructs with the wild type. Bonferoni post hoc range test was used to adjust for multiple comparisons. The GLM Multiple Measures analysis was used to compare activities between promoter constructs measured at different CDCA concentrations.

Construction of ABCB4- and ABCB11 luciferase reporter gene plasmids

In addition to the core promoter upstream of the translation start in exon 2, an alternative promoter upstream of exon –4 was tested using luciferase reporter gene assays.

DNA from three different subjects was used to generate constructs *ABCB4*-B-wt (PromB, core promoter) and *ABCB4*-A-wt (PromA) comprising the reference sequence, and *ABCB4*-A-

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

Typ2A (PromA) containing a combination of 7 variants (IDs: 1, 3, 4, 5, 6, 9, 10; Table 2a). For *ABCB11*, 8 DNA fragments of 2.6 kb of the 5' flanking region of exon 1 comprising 19 polymorphisms from different subjects were generated to obtain 14 different *ABCB11* promoter constructs (Fig. 5). Sequence integrity was confirmed by sequencing.

ABCB4 constructs *ABCB4*-A-wt (2.2 kb), *ABCB4*-A-Typ2A (2.2 kb), *ABCB4*-B-wt (2.8 kb) and *ABCB11* constructs C_1 to C_14 had *Mlu*I and *Bcl*I sites introduced for subsequent cloning into the luciferase reporter gene plasmid pGL3-basic (Promega Catalys AG, Mannheim, Germany).

Luciferase Reporter Gene Assay

HepG2 and human hepatoma (Huh7) cell lines were purchased from ATCC and maintained in RPMI640 (Sigma) supplemented with 10% fetal calf serum (GibcoBRL, Karlsruhe, Germany), 100 U/ml penicilline and 100 mg/ml streptomycin (GibcoBRL, Karlsruhe, Germany). For transactivation assays, cells were grown for 3 days in medium containing 10% charcoal-stripped bovine calf serum and then selected at 90-95% density in 24-well plates. For transient transfection, 1.5 ml of Lipofectamine 2000 reagent (GibcoBRL, Karlsruhe, Germany) and 500-700 ng of plasmid DNA were used per well. Plasmid DNA comprised 450 ng of *ABCB4* construct and 50 ng pSV-beta-galactosidase plasmid or 450 ng *ABCB11* construct, 200 ng *FXR*- and 50 ng pSV-beta-galactosidase plasmid. For *ABCB11*, the cells were treated with up to 200 µM of chenodeoxicholic acid (CDCA). Cells were lysed with passive lysis buffer (PBL; Promega Catalys AG, Mannheim, Germany) 24 hours after transfection. Luciferase activity was quantified with the luciferase assay system (Promega Catalys AG, Mannheim, Germany) using the Lumat LB 9507-2 luminometer (Berthold, Bad Wildbad, Germany). Beta-galactosidase activity was quantified with a high-sensitivity assay (Stratagene, La Jolla, CA) in an UVmax kinetic mircoplate reader (Molecular Devices,

Sunnyvale, CA) at 595 nm. The pGL3-basic plasmid served as control in each separate experiment.

Results:

Overall, we discovered 76 and 86 polymorphic sites in approximately 17 and 13 kb of the *ABCB4 and ABCB11* gene, respectively. The average number of polymorphic sites per kilobase of DNA for *ABCB4* (n=318) and *ABCB11* (n=392) was 4.5 and 6.3 in total, 3.6 (14 in 3.8 kb; 2.1 for missense mutations) and 6.8 (27 in 3.9 kb; 2.5 for missense mutations) in the coding regions, 3.0 (22 in 7.4 kb) and 4.3 (27 in 6.3 kb) in the intronic region, and 7.2 (40 in 5.5 kb) and 10.8 (27 in 2.5 kb) in the 5' UTR. The number of polymorphic sites in the coding region of other ABC transporters was 3.3 (15 in 4.6 kb, n=206) and 5.2 (20 in 3.8 kb, n=494) in MRP3 and MDR1, respectively (Kroetz et al. 2003; Lang et al., 2004). Eight of the 14 (*ABCB4*) and 10 of the 27 (*ABCB11*) variants within the coding regions were non-synonymous mutations (altered the protein sequence), and 9 were not reported in dbSNP or literature previously. As expected, none of the disease-causing PFIC1-3 mutations were detected in our sample set. Genetic variants including their localization and allele frequencies are listed in Tables 2a and 2b.

Polymorphisms in the exons of the ABCB4 and ABCB11 genes

ABCB4: From 19 genetic variants in coding and non-coding exons, 8 included missense mutations, 6 were silent mutations and 5 mutations were located in the untranslated exons -3, -2, -1 and 1. These included three Caucasian variants in exon 6 (c.523A>G; allele frequency 3.2%), exon 16 (c.1954A>G; 7.3%) and exon 26 (c.3296A>G; 1.8%) resulting in amino acid substitutions p.T175A, p.R652G and p.E1099G, respectively. Five rare missense mutations

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

occurred once as single heterozygotes (singletons) in exon 4 (c.261T>C and c.283C>T), exon 10 (c.1099A>G), exon 12 (c.1349A>G) and exon 15 (c.1769G>A) resulting in the amino acid changes p.I367V, p.R590Q, p.D87E, p.P95S, and p.E450G. All amino acid-changing variants of *ABCB4* were predicted to be located in the extracellular and cytoplasmatic region (Fig. 1a). Six SNPs (5 Caucasian, 1 Korean) were detected in the untranslated exons 1, -1, -2 and -3 of *ABCB4*, a region that possibly could influence mRNA stability.

ABCB11: 28 genetic variants were detected in 27 coding and 1 non-coding exons; among these were 10 missense mutations, 17 silent mutations and 1 mutation in the untranslated exon 1. Two Caucasian-specific variants in exon 13 (c.1331T>C; 59.4%) and exon 17 (c.2029A>G; 4.2%) coded for amino acid substitutions p.V444A and p.M677V, one variant detected in exon 16 (c.1846C>G, 2.2%) in the African American population sample resulted in protein sequence alteration p.R616G and the Japanese-specific exon 21 variant c.2594C>T (2.4%) resulted in amino acid substitution p.A865V (Table 2b). Six singletons were detected in exon 8 (c.616A>G → p.I206V), exon 9 (c.851T>C → p.V284 and c.A896G>A → p.R299K), exon 16 (1855A>G → p.T619A), exon 18 (c.2093G>A → p.R698H) and exon 23 (c.2873G>A → p.R958Q). All amino acid polymorphisms of *ABCB11* were predicted to be located in the extracellular region (Fig. 1b).

Genetic variations in the 5' flanking regions and introns

In *ABCB4*, 40 genetic variants were identified in the functional promoter region 2.6 kb and 2.8 kb upstream of exon 2 (21 variants) and exon –4 (19 variants), respectively. In *ABCB11*, 27 genetic variants were detected in the 5' flanking (promoter) region 2.5 kb upstream of exon 1. Furthermore, 22 (*ABCB4*) and 30 (*ABCB11*) genetic variants were detected in adjacent intronic regions.

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

Prediction of functional consequences

Potential consequences of non-synonymous variants in *ABCB4* and *ABCB11* were predicted by means of 5 different computational methods to prioritize further investigations: (1) The type of amino acid change was categorized based upon non-synonymous Grantham values, which provide a measure of chemical similarity. Grantham scores (possible range from 5-215) of less than 50 were classified as non-synonymous conservative, moderately conservative (51-100), moderately radical (101-150), or scores of >151 as radical (Stephens et al., 2001). (2) SIFT (sorting intolerant from tolerant) is an algorithm for predicting functional consequences of amino acid substitutions which assigns scores considering alignments of orthologous sequences. SIFT scores range from 0 to 1: scores <0.05 indicate variant sites in codons for evolutionarily conserved amino acids that are predicted to be deleterious, whereas those ≥0.05 are more likely to be tolerated. (3) PolyPhen (Polymorphism Phenotyping) uses empirically derived rules and computes the absolute value of the difference between profile scores for both variants to predict the likelihood whether a non-synonymous SNP affects protein function or structure. Large differences (>1.5) indicate that the substitution is rarely or never observed in the protein family and therefore more likely to affect the protein. PSIC scores below 0.5 denote benign variants, while PSIC scores between 1.5 and 2 are possibly damaging and PSIC scores above 2 are probably damaging. (4) The amino acid substitution matrix BLOSUM62 applies the same criteria as Cargill et al., (2002) and was used to predict how evolutionarily favorable a non-synonymous SNP is. Scores range from -4 to +3 and substitutions with scores <0 or ≥0 are evolutionarily less or more favorable, respectively. (5) We classified amino acid changes as evolutionarily conserved (EC) or evolutionarily unconserved (EU) based on sequence alignments with two mammalian orthologs (e.g. rat,

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

mouse, and / or rabbit). Laebman et al., (2003) and Shu et al., (2003) showed that substitutions at EC positions are more deleterious than those at EU positions.

The results of these computational analyses are summarized in Table 3. However, prediction results are not totally consistent between the different methods. Two *ABCB4* and two *ABCB11* variants were predominantly predicted to have functional consequences. In particular *ABCB11* p.R616G is suspicious for changing the physicochemical properties of the resulting protein (highest Grantham score of 125), is probably deleterious according to the low SIFT value (0.01), is supposed to affect protein function due to its high PSIC score, and is evolutionarily less favorable considering the negative BLOSUM62 value, although sequence alignments indicated that p.R616G is located in a probably less deterious evolutionarily unconserved region. Similarly, *ABCB4* p.E1009G and *ABCB11* p.T619A have more radical Grantham scores, are less tolerated substitutions (SIFT) with possibly damaging effect (PSIC) and are evolutionary less acceptable but located in an evolutionary unconserved region. In contrast, *ABCB4* p.R652G is a more common variant, which changes an evolutionary conserved amino acid to an evolutionary less acceptable one with probably altered physicochemical properties. On the other hand, SIFT and PSIC scores of this variant do not confirm functional consequences of *ABCB4* p.R652G.

Furthermore, reference and variant genomic sequences of *ABCB4* and *ABCB11* were used to predict potential splice site variants (www.fruitfly.org/seq_tools/splice.html). None of the detected intronic mutations were identified to alter consensus sequences of existing splice sites.

Ethnic specifity of SNPs

ABCB4 was investigated in Caucasian, Japanese, and Korean populations, whereas *ABCB11* was investigated in Caucasians, Japanese and AfricanAmericans. Twenty nine SNPs occurred

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

in all three populations in *ABCB4* and 15 SNPs occurred in all three populations in *ABCB11*. Most *ABCB4* and *ABCB11* variants were population-specific and either their occurrence or their population frequency varied between different ethnic groups. Not surprisingly, the largest genetic diversity was detected in samples of AfricanAmerican origin (*ABCB11*). They had the highest number and proportion (30/54) of population-specific alleles, compared to 11/37 in the Caucasian sample and 15/32 in the Japanese and compared to 14/47, 11/47, and 13/47 population-specific *ABCB4* SNPs in Caucasian, Japanese, and Koreans, respectively.

Few population-specific SNPs were observed at higher frequencies. In *ABCB4*, only the Caucasian population had 1/14 population-specific variants at a frequency ≥5%. In contrast, in *ABCB11* the AfricanAmerican population sample had 6/30 at frequencies ≥5%, while Caucasians had no population-specfic alleles at ≥5%. 43% of the AfricanAmerican-specific mutations (*ABCB11*), 72% of Caucasian-specific and 82% of Asian-specific variants (*ABCB11* and *ABCB4*) were singletons.

Rare variants are more likely to be recently derived than are the common variants and are, therefore, more likely to be population-specific. Hence, they are sensitive indicators for the relationships among populations. As expected, our Japanese sample shared more rare variants (<5%) with Korean samples than with Caucasian or African American samples. Most *ABCB4* and *ABCB11* variants that were found in all populations had allele frequencies ≥5%.

ABCB4 p.R652G was the only protein-altering variant with high allele frequency in all groups (7.2% in Caucasian, 1.4% in Japanese, and 2.3% in Korean). *ABCB4* p.T175A (3.2%) and p.E1099G (1.8%) were only present in the Caucasian sample. The other 5 missens mutations were singletons. The most common *ABCB11* protein-altering polymorphism was p.V444A, which was frequently observed in all groups (*ABCB11* p.M677V was present in both the Caucasian (4.2%,) and the AfricanAmerican (14%) population sample and p.A865V was only found in Japanese samples. The other 6 non-synonymous mutations were singletons.

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

Population Genetic Analysis

Nucleotide diversity provides a measure of genetic variation that is normalized by the sample size. We estimated two measures of nucleotide diversity, the average heterozygosity per site (π), and the population mutation parameter (θ). In addition, Tajima´s *D* was calculated to detect deviations from the neutral mutation model (Tajima 1989). These parameters were estimated for all variable sites with less than 15% of missing data for various gene regions (coding region, non-coding region, exon-intron boundaries and 5'UTR) as well as for various sites within the coding region (synonymous and non-synonymous sites) and separately for each population (Table 4). It should be noted that only 36 of 54 segregating sites from African Americans were included in the calculations and the estimates may therefore not reflect the actual population values. The diversity estimates were similar across ethnic groups and within the range previously reported for other genes in various ethnic groups (Cargill et al., 1999; Halushka et al., 1999; Glatt et al., 2001; Laebman et al., 2002). The estimates of θ ($\times 10^{-4}$) ranged from 4.39 to 4.98 (*ABCB4*) and 4.53 to 5.52 (*ABCB11*) and those of π ($\times 10^{-4}$) ranged from 2.79 to 4.31 (*ABCB4*) and 3.63 to 5.28 (*ABCB11*) for the entire sequenced region (Table 4). Furthermore, θ values were higher compared to the corresponding π values in 5 of 6 population groups and Tajima´s *D* values were positive only for *ABCB11* in the Japanese (0.36), and negative in all other population samples and for *ABCB4*. However, neither estimate was statistically significant. The highest θ and π values were estimated in the 5´UTR in both genes in all population samples.

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

Linkage Disequilibrium of genetic variants in the ABCB11 and ABCB4 gene

Linkage disequilibrium (LD) was evaluated for the Caucasian, African American, Japanese, and Korean sample populations. Significant disequilibrium between site pairs as calculated by Fisher's exact test and blocks of LD calculated using the confidence-bound method (Gabriel et al., 2005) are shown in figures 2a and 2b. The D' values between all site pairs are generally much higher then the r^2 values with a large proportion of D' values equal to 1.0 or -1.0 (maximum disequilibrium). Several groups of SNPs in *ABCB4* and *ABCB11* were in tight LD (most D' values are 0.7 to 1.0 and -1.0) with each other and there is evidence for single block structures. In general the largest LD blocks and the highest number of alleles within the individual LD blocks were found in *ABCB4* compared to *ABCB11*.

In *ABCB4*, 3 groups of genetic variants appeared to be in strong LD. The largest group comprised 19 SNPs (IDs 1,3,4,5,6,9,10,14,16,17,19,24,26,28,30,33, and 37) in all population samples and additionally ID 41 in Japanese and Korean, or ID 42 in the Caucasian population sample. Linkage disequilibrium between these SNPs occurred across an 80 kb distance spanning almost the entire *ABCB4* gene. Other groups of variable sites in strong LD were observed in Caucasians between IDs 2,6,11,12,20,21,31,32,35,36,38,39 and between the 8 variable sites 20,21,31,32,35,36,38,39 in Japanese and Koreans. The highest number of alleles (n= 11-14) was found in the LD block1 of all populations consisting mainly of promoter SNPs with IDs 1-23. In Caucasian the LD blocks3 and 4 were the largest containing exon/intron alleles.

In *ABCB11*, the bulk of significant linkage disequilibrium and largest LD blocks occurred in Caucasians in 3 groups of segregating sites, between the 9 SNPs (IDs 1,2,3,4,7,8,39,40,41), the 7 SNPs (IDs 29,30,31,39,40,41,43), and between the 4 SNPs (IDs 44,47,51,53). In African Americans, significant LD was observed between the 9 variable sites 6,7,10,18,20,23,39,-40,41 and 13,15,20,25,25,31,33,35,38 as well as between the 4 variable sites 45,46,48,50 and

3 variable sites 44,51, and 53. In Japanese significant LD was detected among the 6 segregating sites 18,20,23,39,40,41, the 5 variable sites 1,2,3,4,8, the 4 variables sites 32,39,40,41 and the three variable sites 44,51,53.

Haplotype structure

The combination of alleles present at each segregating site in each individual was computed by assigning a specific pair of haplotypes to each individual, as well as a score reflecting the confidence in that assignment. Haplotypes were statistically inferred by PHASE separately for each population group. Resulting haplotypes are arranged according to sequence similarity in table's 5a and 5b. The *ABCB11* and *ABCB4* reference sequences are composites assembled from different sources as given in the method section, and as expected; they were not encountered in our study population.

ABCB4 haplotype structure

Altogether, 72 haplotypes, 29 in Caucasians, 22 in Japanese, and 21 in Koreans were identified with high confidence. Among them were 8 common haplotypes, *ABCB4*_11, *ABCB4*_12, *ABCB4*_16, *ABCB4*_20, *ABCB4*_23, *ABCB4*_26, *ABCB4*_40 and *ABCB4*_45 accounted for 72% of the 318 chromosomes. Both, number and frequency distribution of haplotypes were subject to great interethnic variability. Twenty haplotypes (28%) were specifically detected only in the Caucasian population sample, compared to 10 (14%) Japanese-specific and 11 (15%) Korean-specific haplotypes. The ethnic distribution of the 8 most common *ABCB4* haplotypes is illustrated in figure 3a. *ABCB4*_12 and *ABCB4*_20 were the major alleles in all populations and detected at similar frequencies (Caucasian: 32.8% and 18.0% Japanese: 36.2% and 11.7% and Korean: 25.0% and 18.8%), while *ABCB4*_45 was a dominant haplotype in both Asian groups. *ABCB4*_40 and *ABCB4*_45 were specific to

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

Japanese and Korean individuals and *ABCB4*_16 and *ABCB4*_23 were specific to the Caucasian population.

The reference sequence (GenBank accession number NM_000443) was not found in the entire sample set. However, except for the fact that the random mutation variant me178 is contained, haplotype-*ABCB4*_15 is identical to the reference sequence and was found in 8% of Caucasians. The two major allelic variants *ABCB4*_12 and *ABCB4*_20 carried 5 SNPs including 1 promoter, 3 intronic, and 1 synonymous. Of the eight common haplotypes, three (*ABCB4*_11 *ABCB4*_12 *ABCB4*_16) contained one coding SNP, three (*ABCB4*_23, *ABCB4*_26, *ABCB4*_45), had two coding SNPs and none included a non-synonymous variant.

ABCB11 haplotype structure

Altogether, the 53 variant sites segregated as 38 (CA), 41 (AA) and 30 (JA) distinct haplotypes, which were identified with high confidence. Only 38 haplotypes were present in 3 or more chromosomes. The percentage of chromosomes in the entire population that could be assigned to one of these 38 common haplotypes was 64.1% and differed between groups (85.6%, 43.6% and 64.6% for Caucasian, African American and Japanese individuals, respectively). The 10 most common haplotypes, *ABCB11*_3, *ABCB11*_4, *ABCB11*_5, *ABCB11*_6, *ABCB11*_11, *ABCB11*_26, *ABCB11*_28, *ABCB11*_37, *ABCB11*_38 and *ABCB11*_65 accounted for 57.9%, 23.4% and 45.8% of the 202 Caucasian, 94 African American and 96 Japanese chromosomes, respectively (Fig. 3b).

*ABCB11*_3, *ABCB11*_6, and *ABCB11*_28 were the major alleles in Caucasians while *ABCB11*_3, *ABCB11*_26 and *ABCB11*_37 were dominant in the Japanese and *ABCB11*_11 and *ABCB11*_65 are predominant in the African American population, respectively. The cDNA sequence with the GenBank accession number NM_003742 was defined as reference. With respect to the reference sequence *ABCB11*_3, *ABCB11*_4, *ABCB11*_5, *ABCB11*_6,

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

*ABCB11*_11 *ABCB11*_37 and *ABCB11*_38 contain one non-synonymous cSNP (c.1331T>C) in exon 13 (p.V444A) and three to four linked intronic SNPs in intron 9, 13 and 14 and 20 as a common denominator. Furthermore, *ABCB11*_4, *ABCB11*_6 and *ABCB11*_38 contain two additional intronic and synonymous SNPs in intron 28 and exon 24 (p.A804A). Furthermore, a number of population-specific haplotypes were detected (25 in Caucasians 33 in African Americans and 22 in Japanese). Most of these unique haplotypes differed only in their promoter or intronic sequence from the common alleles. As the synonymous polymorphism in exon 6 (p.I134I) is specific for the African American population, all haplotypes containing this SNP are specific for this ethnic group (12 haplotypes). Similarly, all haplotypes containing the Japanese-specific variant in exon 4 (p.D36D) and exon 9 (p.Y269Y) are specific for the Japanese population (12 haplotypes).

ABCB4 and ABCB11 DNA promoter activity

Reporter gene activity was measured for all three constructs in HepG2 and Huh-7 cells after transfection of luciferase gene under the control of the 5' flanking region of *ABCB4*. Neither the *ABCB4*-A-wt nor the *ABCB4*-A-Typ2A constructs showed promoter activity in either cell line, as compared to the strong induction of luciferase activity that was conferred by construct *ABCB4*-B-wt (Fig. 4).

The luciferase activity of 13 *ABCB11* promoter constructs comprising mutations at 19 segregating sites was compared with wild type in HepG2 cells (Fig. 5). In two independent experiments, promoter constructs C_2 and C_3 showed approximately 50% lower activity than wild type ($p<0.0001$ and $p=0.043$ for C_2 and $p<0.0001$ and $p=0.045$ for C_3; GML Univariate procedure with Bonferoni post hoc test). C_2 and C_3 differ from wild type at 4 positions and additionally, C_3 differs from C_2 and wild type at position g.-15150. Since the latter variant was present in a number of promoter constructs without showing any effect, only

C_2 was investigated further. To confirm results, luciferase activity of C_2 was investigated at various CDCA concentrations (Fig. 6). At all CDCA concentrations, C_2 demonstrated a statistical significantly lower activity compared to the wild type construct (p=0.0072; GML Repeated Measures analysis).

Discussion:

We evaluated genetic variability, linkage disequilibrium and haplotype profiles of hepatic efflux transporter genes *ABCB4* and *ABCB11* in 292 individuals of different ethnic populations.

Genetic variation in *ABCB4* and *ABCB11* has recently been investigated with respect to their potential pathogenetic role in primary biliary cirrhosis (PBC) and sclerosing cholangitis (PSC). In these case-control studies our healthy Caucasian population sample served as control group (Pauli-Magnus et al. 2004a,b). In the study at hand, however, *ABCB4* and *ABCB11* genotyping data, including those obtained from Caucasians, have been computed independently. Furthermore, *ABCB4* haplotype analysis considered 13 Prom variants. Our findings with respect to haplotype structures and linkage disequilibrium correspond essentially with those reported from Pauli-Magnus. Not surprisingly, this applies also for *ABCB4*. Most of the *ABCB4* haplotypes from Pauli-Magnus were inferred in our analysis as well. PromA variants clustered in one block resulting in 18 additional haplotypes (MDR3_1, 2, 3, 23, 24, 25, 26, 27, 28, 29, 43, 45, 47, 48, 49, 67, 68, 69).

A major goal was to establish potential hereditary markers as a prerequisite for assessing interindividual genetic variability of cholestatic liver injury such as DIC or ICP. More than 30 mutations in *ABCB4* and *ABCB11* have been associated with lack or low level of protein expression, or cholestatic liver injury such as PFIC2, PFIC3, DIC, or ICP (Pauli-Magnus and

Meier, 2003), although the causative role of most of the mutations for etiology remains to be established. Not surprisingly, most of these rare variants were not detected in our samples. Only two variants from our study had been related to cholestatic disease earlier, *ABCB4* c.523A>G (p.T175A), the second most prevalent non-synonymous change in Caucasians (Rosmorduc et al., 2001) and the most prevalent non-synonymous variant *ABCB4* c.1954G>A (p.R652G), which was present in all of our population samples (Jacquemin et al., 2001). In addition to these known genetic variants, 13 *ABCB4* and 23 *ABCB11* coding variants, among them 3 and 6 missens mutations, have been described for the first time. The functional relevance of these novel variants is unknown. While their absence in various cholestatic disease and presence in healthy volunteers may doubt a significant contribution to risk for more common cholestatic diseases such as PBC and PSC, a role of these variants for DIC or ICP and a contribution of less common mutations, which have been detected only as heterozygots, to familial cholestatic disease can not be excluded familial diseases such as PFIC follow a recessive trait caused by rare mutants. Homozygous carriers of the mutation are affected, whereas heterozygous carriers appear phenotypically normal. Furthermore, a mutation may have suffice bile salt excretion under normal conditions, but combined with another mutation or in circumstances such as systemic infection, pregnancy, or drug intake lead to clinical symptoms. For instance, the MDR3 variant p.R652G was present in PFIC3 patients and healthy subjects and hence can alone not be sufficient to cause cholestasis. It was suggested that p.R652G had no or mild consequences under normal conditions, but resulted in cholestasis during pregnancy (Jacquemin et al., 2001). This view is further supported by the detection of compound heterozygosity. A splicing mutation (+3)A>C (intron 4) combined with a frameshift mutation in exon 22 (p.K930X) resulted in a PFIC2 phenotype and two nonsynonymous variants in exon 9 (p. E297G) and in exon 12 (p.R432T) were encountered in the patient exhibiting a BRIC (benign recurrent intrahepatic cholestasis) phenotype. Single

occurrence of these mutations did not confer any disease risk (Noel et al., 2005). Family studies are required to identify rare mutations that are causative for hereditary recessive disorders, whereas carefully designed, sufficiently powered case-control studies with well-defined cholestatic phenotypes (e.g. ICP or DIC patients) and matched healthy controls could establish, whether the detected *ABCB4* and *ABCB11* genetic variants are of any clinical significance.

Our study provides the common genetic variability in various ethnic populations to facilitate such case-control studies in the future. In the absence of time-consuming and laborative *in-vitro* assays for prediction of a possible functional role of nonsynonymous variants, can computational tools serve as a guidance to prioritize those genetic variants for further analysis, which are most likely functional effective. We combined several computational tools that categorize the type of amino acid change based on physicochemical considerations and grade of evolutionary sequence conservation (Grantham, SIFT, PolyPhen, BLOSUM62, EC/EU). The validity of this approach has been shown recently for the organic cation transporter OCT1 (Shu et al., 2003). Furthermore, disease-causing mutations were more prevalent at evolutionarily conserved sites (Miller and Kumar, 2001). Although experimentally not validated, we conclude that the amino acid changes p.R590Q, p.E1099G (*ABCB4*), p.R616G, and p.T619A (*ABCB11*) are the strongest candidates to alter protein function and subsequently biliary excretion (Table 3).

Population genetic analysis resulted in negative Tajima´s *D* values, suggesting that *ABCB4* and *ABCB11* are under some selective pressure against changes in the protein. Compared to the closely related xenobiotic transporter *ABCB1* (Kroetz et al., 2003), *ABCB4* and *ABCB11* are genetically less diverse and display only few protein-altering changes and particularly few at higher frequency. Possibly, *ABCB4* and *ABCB11* are functionally more constrained, reflecting their important endogenous role. Unlike *ABCB1* with its primary role in xenobiotics

detoxification, *ABCB4* and *ABCB11*, while being involved in disposition of xenobiotics to some extent, primarily maintain homeostasis of endogenous cholephilic compounds. Moreover, protein-altering variants occurred at lower frequency than synonymous changes. Only 1 out of 8 *ABCB4* and 2 out of 10 *ABCB11* amino acid changes occurred at ≥5% compared to half of the synonymous sites (3 of 6 sites in *ABCB4* and 9 of 18 sites in *ABCB11*). This finding is compliant with the hypothesis that deleterious mutations, which are more likely to be at non-synonymous sites, are kept at low frequency and spread less easily to multiple populations (Fay et al, 2001).

Haplotype-based approaches taking into consideration the combination of SNPs, which ultimately represent the functional unit of the gene, are particularly useful to correlate cellular and clinical phenotypes with a specific gene. They have been proven very useful in small population samples even when correlation analysis with SNPs fail (Drysdale et al., 2000). Moreover, attempts to draw associations between phenotypes and genetic variations are more likely to succeed when the SNPs used in such studies have been confirmed to be in linkage disequilibrium by methods such as haplotyping. Thus, understanding the haplotype structure and to what extent diversity and distribution of haplotypes vary across populations is important for designing and interpreting genotype-phenotype association studies. Our haplotypes were statistically inferred without experimental validation of the molecular linkage of SNPs. A few SNPs that do not fit the model could be overlooked and analyzing a large genomic region is likely to have a certain error rate deteriorating the precision of haplotype prediction. This is a potential source for errors, although it is unlikely to affect the result of association analyses, which are based on more common haplotypes. In contrast to other ABC transporters such as *ABCB1* (Kroetz et al. 2003), *ABCB4* and *ABCB11* had no predominant haplotype. Instead, there were multiple haplotypes, most of which are observed in several populations, that accounted for a large fraction of genomic variability. The 2 most common

ABCB4 haplotypes were observed in 25% to 36% and 12% to 19% of all populations, whereas the 4 most common *ABCB11* haplotypes occurred in approximately 15% of distinct populations (Fig. 3a and 3b). There was distinct interethnic variability in the total number and the pattern of population-specific haplotypes and haplotypes that were observed in all populations showed ethnically distinct population frequencies. Haplotype-based association studies to predict MDR3 or BSEP phenotypes need to carefully consider this ethnic population substructure.Smit et al., (1995) characterized 3 kb of the 5' flanking region of *ABCB4* upstream of the translation start in exon 2 and located the core promoter. Genetic variants of this core promoter have been tested for differential activity in luciferase assays (Pauli-Magnus et al., 2004). Additionally, an aberrant 6-kb *ABCB4* cDNA (accession number: Z35284) containing 4 untranslated exons (-1, -2, -3, -4) indicated the presence of an alternative transcription start upstream of exon -4 (Smit et al., 1995). In our luciferase reporter gene assays, plasmid constructs containing the alternative promoter of *ABCB4* (PromA) showed no significant activity in the liver cell lines HepG2 and Huh-7. Although we could not detect promoter activity, the function of this putative regulatory element remains unclear. It may only be active in entity with the entire *ABCB4* gene, in conjunction with the core promoter in a tissue specific manner, or only under certain physiological conditions. From 13 *ABCB11* promoter haplotypes comprising 19 variants, two (C_2, C_3) altered promoter activity compared to wildtype. At least one of variants 13, 15, or 83 is responsible for the significant decrease of promoter activity. A very interesting candidate is the short allele of a variable length [T]n polymorphism (9 to 13 T-repeats, IDs 83 to 86, respectively), which is common in all populations (Tables 2b and 5). Promoter nucleotide-repeats have been previously reported to be required for the binding of transcription factors with the length of the T-stretch influencing transcription (Beutler et al., 1998).

In conclusion, our findings give a comprehensive overview on genetic variability, haplotype structure, ethnic diversity, and linkage disequilibrium of two important hepatic transporters. Based on computational predictions it seems likely that some genetic variants may contribute to interindividual variability in MDR3 and BSEP function. In contrast to xenobiotic transporters such as MDR1, which control the access of substrates to pharmacological sanctuaries, changes in MDR3 and BSEP transport function may have their greatest impact on occurrence, pattern and prognosis of cholestatic liver injury. Furthermore, they may modulate the individual sensitivity to drug-mediated inhibition of MDR3 and BSEP transport and consequently the individual sensitivity to DIC. Further *in-vitro* approaches and clinical studies need to clarify the potential functional and clinical consequences of the variants identified in this study and to whether these contribute to cholestatic disease or to altered sensitivity for drug-mediated inhibition of MDR3 or BSEP transport.

References:

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21:** 263-265.

Beutler E, Gelbart T, Demina A (1998) Racial variability in the UDP-glucuronosyltransferase 1 (UGT1A1) promoter: a balanced polymorphism for regulation of bilirubin metabolism? *Proc Natl Acad Sci U S A* **95:**8170-8174.

Borst P, Elferink RO (2002) Mammalian ABC transporters in health and disease. *Annu Rev Biochem* **71:**537-592.

Cargill M, Altschuler D, Ireland J, Sklar P, Ardilie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* **22:** 231-238.

Drysdale CM, McGraw DW, Stack CB, stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G, Ligget SB (2000) Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. *Proc Natl Acad Sci USA* **97:**10483-10488.

Fay JC, Wyckoff GJ, Wu CI (2001) Positive and negative selection on the human genome. *Genetics* **158:**1227-1234.

Funk C, Ponelle C, Scheuermann G, Pantze M (2001) Cholestatic potential of troglitazone as a possible factor contributing to troglitazone-induced hepatotoxicity: in vivo and in vitro interaction at the canalicular bile salt export pump (Bsep) in the rat. *Mol Pharmacol.* **59:**627-635.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. (2002) The structure of haplotype blocks in the human genome. *Science* **296:** 2225-2229.

Glatt CE, De Young, Delgado S, Service SK, Giacomini KM, Edwardy RH, Risch N, Freimer NB (2001) Screening a large reference samle to identify very low frequency sequence variants: comparison between two genes. *Nat. Genet.* **27:**435-438.

Grantham R (1974) Amino Acid difference formula to help explain protein evolution. *Science* **185:** 862-864.

Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat. Genet.* **22:**239-462.

Jacquemin E, De Vree JM, Cresteil D, Sokal EM, Sturm E, Dumont M, et al. (2001) The wide spectrum of multidrug resistance 3 deficiency: from neonatal cholestasis to cirrhosis of adulthood. Gastroenterology **120:**1448-1458.

Kerb R (2006). Implications of genetic polymorphisms in drug transporters for pharmacotherapy. *Cancer Lett* **234**:4-33.

Kostrubsky VE, Strom SC, Hanson J, Urda E, Rose K, Burliegh J, Zocharski P, Cai H, Sinclair JF, Sahi J (2003) Evaluation of hepatotoxic potential of drugs by inhibition of bile-acid transport in cultured primary human hepatocytes and intact rats. *Toxicol Sci.* **76:**220-228.

Kroetz DL, Pauli-Magnus C, Hodges LM, Huang CC, Kawamoto M, Johns SJ, et al (2003) Sequence diversity and haplotype structure in the human ABCB1 (MDR1, multidrug resistance transporter) gene. *Pharmacogenetics* **13:**481-494.

Kullak-Ublick GA, Stieger B, Meier PJ (2004) Enterohepatic bile salt transporters in normal physiology and in liver disease. *Gastroenterology* **126:**322-342.

Laebman MK, Huang CC, Kawamoto M, Johns SS, Stryke D, Ferrin TE, De Young J, Taylor T, Clark AG, Herskowitz I and Giacomini KM on behalf of the Pharmacogenetics of Membrane Transporters Investigators (2002) Polymorphisms in ahuman kidney xenobiotic transporter, OCT2, exhibit altered function. *Pharmacogenetics* **12:** 395-405.

Laebman MK, Huang CC, DeYoung J, Carlson EJ, Taylor TR, de la Cruz M, Johns SJ, Stryke D, Kawamoto M, Urban TJ, Kroetz DL, Ferrin TE, Clark AG, Risch N, Herskowitz I, Giacomini KM and Pharmacogenetics of Membrane Transporters Investigators (2003) Natural variation in human membrane transporter genes reveals evolutionary and functional constraints. *Proc. Natl. Acad. Sci. USA* **100:**5896 - 5901.

Lang T, Hitzl M, Burk O, Mornhinweg E, Keil A, Kerb R, Klein K, Zanger UM, Eichelbaum M, Fromm MF (2004) Genetic polymorphisms in the multidrug resistance protein 3 (ABCC3) gene and relationship to its mRNA and protein expression in human liver. *Pharmacogenetics* **14:** 1-10.

Miller MP, Kumar S (2001) Understanding human disease mutations through the use of interspecific genetic variation. *Hum Mol Genet* **10:**2319-2328.

Noe J, Kullak-Ublick GA, Jochum W, Stieger B, Kerb R, Haberl M, Mullhaupt B, Meier PJ, Pauli-Magnus C (2005). Impaired expression and function of the bile salt export pump due to three novel ABCB11 mutations in intrahepatic cholestasis. *J Hepatol* **43:**536-43

Pauli-Magnus C, Meier PJ (2003) Pharmacogenetics of hepatocellular transporters. *Pharmacogenetics* **13:**189-198.

Pauli-Magnus C, Lang T, Meier Y, Zodan-Marin T, Jung D, Breymann C, Zimmermann R, Kenngott S, Beuers U, Reichel C, Kerb R, Penger A, Meier PJ, Kullak-Ublick GA (2004) Sequence analysis of bile salt export pump (ABCB11) and multidrug resistance p-glycoprotein 3 (ABCB4, MDR3) in patients with intrahepatic cholestasis of pregnancy. *Pharmacogenetics* **14:**91-102.

Pauli-Magnus C, Kerb R, Fattinger K, Anwald Stieger B, Fattinger K, Madon J, Kullak-Ublick GA, Meier PJ (2004) BSEP and MDR3 haplotype structure in healthy Caucasians, primary biliary cirrhosis and primary sclerosing cholangitis. *Hepatology* **39:**779-791.

Rosmorduc O, Hermelin B, Poupon R (2001) MDR3 gene defect in adults with symptomatic intrahepatic and gallbladder cholesterol cholelithiasis. *Gastroenterology* **120:**1459-1467.

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analysis by the coalescent and other methods. *Bioinformatics* **19:**2496-2497.

Saito S, Iida A, Sekine A, Miura Y, Ogawa C, Kawauchi S, et al. (2002) Three hundred twenty-six genetic variations in genes encoding nine members of ATP-binding cassette, subfamily B (ABCB/MDR/TAP), in the Japanese population. *J Hum Genet* **47:**38-50.

Schneider S, Kueffler JM, Roessli D, Excoffier L (2000) *Arlequin ver 2.000: A software for population genetics data analysis.Geneva*. Switzerland: Genetics and Biometry Laboratory, University of Geneva.

Shu Y, Laebman MK, Feng B, Mangravite LM, Huang CC, Stryke D, Kawamoto M, Johns SJ, DeYoung J, Carlson E, Ferrin TE, Herskowitz I, Giacomini KM, and Pharmacogenetics of Membrane Transporter Investigators (2003) Evolutionary conservation predicts function of variants of the human organic cation transporter, OCT1. *Proc. Natl. Acad. Sci. USA* **100:**5902-5907.

Smit JJ, Mol CA, van Deemter L, Wagenaar E, Schinkel AH, Borst P (1995) Characterization of the promoter region of the human MDR3 P- glycoprotein gene. *Biochim Biophys Acta* **1261:**44-56.

Smith AJ, van Helvoort A, van Meer G, Szabo K, Welker E, Szakacs G, et al. (2000) MDR3 P-glycoprotein, a phosphatidylcholine translocase, transports several cytotoxic drugs and directly interacts with drugs as judged by interference with nucleotide trapping. *J Biol Chem* **275:**23530-23539.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68:**978-989.

Stieger B, Fattinger K, Madon J, Kullak-Ublick GA, Meier PJ (2000) Drug- and estrogen-induced cholestasis through inhibition of the hepatocellular bile salt export pump (Bsep) of rat liver. *Gastroenterology* **118:**422-430.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585-595.

Footnotes

Supported by grants 01GG9846 and 01GG9848 from the German Federal Ministry for Education and Science (BMBF).

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

Figures

Fig. 1: MDR3 (a) and BSEP (b) secondary structures and variant locations. The two-dimensional transmembrane topologies of the transporters MDR3 and BSEP were predicted using TOPO (S.J. Johns and R.C. Speth, transmembrane protein display software, www.sacs.ucsf.edu/TOPO/topo.html, unpublished). Common variants are unframed, healthy Caucasian-specific variants are shown in grey, African American-specific variants in black, Japanese-specific variants in white and Korean-specific variants with diagonal pattern.

Fig. 2: *ABCB4* (a) and *ABCB11* (b) pairwise linkage disequilibrium. The triangle plots show pairwise linkage disequilibrium for each population. Statistical significance is indicated by different colours and was assessed by a Fisher´s exact test. White square: not significant, light grey square: $0.05 > p > 0.01$, dark grey square: $0.01 > p > 0.001$, and black square: $p < 0.001$, white square with dot in the center: site pairs that lack the power to detect a significant association. Triangles indicate LD blocks identified by Haploview.

Fig. 3: Ethnic distribution of the most common *ABCB4* (a) and *ABCB11* (b) haplotypes.

Fig. 4: Huh7 and HepG2 cells were transfected with the indicated constructs. Promoter activity is shown as the ratio of luciferase to beta-galactosidase. Results are expressed as the mean ± 1 S.E. of 4-6 transfections. *ABCB4* reporter constructs (*ABCB4*-B-wt) showed a 100-fold higher luciferase activity compared to the empty pGL3 basic vector (basic) in both cell lines. *ABCB4* reporter gene constructs containing *ABCB4*-A-wt and polymorphisms in the

ABCB4 gene (*ABCB4*-A-Typ2A) showed no luciferase activity compared to the empty pGL3 basic vector (basic) in both cell lines. Promoter A sequence did not confer basal luciferase activity.

Fig. 5: Haplotypes of *ABCB11* promoter clones tested for luciferase activity using reporter gene assays. A dot indicates that the nucleotide at this position is identical with the wildtype.

Fig. 6: Promoter luciferase activity of *ABCB11* promoter construct C_2 (◊) versus wildype (O) at various concentrations of CDCA measured in HepG2 cells. Results are expressed as mean ± SD of 3 experiments. Promoter activity of C_2 was statistical significantly lower compared to wildtype (p=0.0072; GML Repeated Measures analysis).

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

Table 1a: PCR amplicon information for *ABCB4*

Fragment name	Primer designation	5´ - 3´primer sequence	Product size (bp)
PromA7	ME_PA7f	CAAGTTCCGACAACTGGGATG	498
	ME_PA7r	TGCAATACTCAGAGGACCACTAAAA	
PromA6	ME_PA6f2	GACAATCAGTAGATCCTTCTATTCA	
	ME_PA6r1	AGTTCCAAGGTCACATTGCCTAATG	637
PromA5	ME_PA5f	ATCCATTTCATCGTTGCCTTG	
	ME_PA5r	CCCTTCTCATTGCTTGACTCAGA	441
PromA4	ME_PA4f	TGCTGGGCCGAATTAGATGAA	
	ME_PA4r	CAGAGAATAACAAATTCCATCTACC	541
PromA3	ME_PA3f	TCCATCATAGCACCCTCTTTAGGAA	
	ME_PA3r	AAGGACCTGGGCTAAGGTTTTAGGA	554
PromA2	ME_PA2f	CCTTGAACATGCTCCTTGTTTTGA	
	ME_PA2r	AGCACTCCACCTCAGGGTGTTT	510
PromA1	ME_PA1f	GGAACTGAGAGTCCTGAGGGTACA	
	ME_PA1r	GAATCTGACCTTTGCCCTGCAC	499
PromB8	ME_PB8f	GGCAATGTAAAAGTTAGGGGTGAA	
	ME_PB8r	TGCAGTATCACTGAGAATTCCATC	456
PromB7+Exon-3	ME_PB7f	ACAGCCTGAGTGTAATGGAGAAA	
	ME_PB7r	CAAATTCCTTCCCTTCACTGGT	506
PromB6+Exon-3	ME_PB6f	CCAGTCACCACAAGCCTTCTTT	
	ME_PB6r	TTTCCCATGTACCAGCTGTCTG	501
PromB5+Exon-2	ME_PB5f	GGCCTCGCCTGAACTTTTACTA	
	ME_PB5r	TGTGCTCATTTCTCTAGGCCTGT	502
PromB4	ME_PB4f	TCGCAATAACCTGATAAGGGAGA	
	ME_PB4r	CTGTTGGTGCCCTCTTAGCACT	482
PromB3+Exon-1	ME_PB3f	AGCCTCAGGCAGGACTTCAC	
	ME_PB3r	AAGCCCCTGCAGATGACTGT	389
PromB2	ME_PB2f	ATTGCTCTCTCTCGGTCCTG	
	ME_PB2r	CAGAGGGGCCTGGACTTTG	440
PromB1+Exon1	ME_1f	GGCTGCAACGGTAGGCGTTT	
	ME_1r	GGCGTGTAACGGAAAAGCCAGT	434
Exon2	ME_2f	GCGAGGTTCGAGGTGAGAGA	
	ME_2r	AACCGGATGCAAGACCCTTC	404
Exon3	ME_3f1	CTCTGTAAGGCAGGCGACCA	
	ME_3r1	CTCAAGCAACCCTCCCACCT	453
Exon4	ME_4f	GAAGGCCTCCTTTTCTAAGACATTC	
	ME_4r	TGGAGTCAACCAGATATCCAAATC	416
Exon5	ME_5f	CAGTACCTAAACCCTTGGGCTCT	
	ME_5r	GGGCCATGATGTGTGCATCT	432
Exon6	ME_6f	TGAGATGGTGCCACTGCACTC	
	ME_6r	ACATGGCTGCCAGATGATCG	450
Exon7	ME_7f	CTGAAAGAGGCTTGCAGTCAGTG	
	ME_7r	CACCCAGCCTGTGACATTTTGAA	488
Exon8	ME_8f	CCAGAACACTGGCATTTGCTAC	

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

	ME_8r	TAGCCATCAGTAAAGGGTGCTTA	415
Exon9	ME 9f	CGAGTGTGACTCGGACTATGG	
	ME_9r	GAAAGTCAGCTCTGCCCACT	402
Exon10	ME 10f	CCTTGCAAATGTTGCTCTTCCAC	
	ME_10r	GCTCAAAGACTTCTTTTGGCACTA	309
Exon11	ME 11f	CCAGGTCCTATTTTTGGAATTTGCTG	
	ME_11r	AACCCCCAAAGGAAAAGGCACATA	322
Exon12	ME 12f	GTGCCTTTTCCTTTGGGGGTTA	
	ME_12r	TGAAACCAGCCCAAGGGTGT	423
Exon13	ME 13f	GGATGTTTTTCATGAATGGTCCTG	
	ME_13r	GCAAAGTTGGACAATCTTGCATC	403
Exon14	ME 14f	CCAACTTTGCTAGATAAGTCCCCT	
	ME_14r	CGGAAGCACTGGCAAGAATC	500
Exon15	ME 15f	ACAAACAGCATTCATCCAAGTGC	
	ME_15r	TGGCACTTTCCCCTATTTTCTCA	450
Exon16	ME 16f	CATCCATTTGGAGACACACACAC	
	ME_16r	GTAGCAGTCATCTGTGCCTGAAA	352
Exon17	ME 17f	TGCACCTAGTTTGCGAGTCC	
	ME_17r	CAGAATGGAGCCAGTCAGTG	350
Exon18	ME 18f	CTGGTGAAGGTTGCACAACAATG	
	ME_18r	CAGCAGAGCCTTATGCCAATCA	468
Exon19	ME 19f	TGAGGATCCGTGGCAACTGTA	
	ME_19r	CAGGACTCTGCCCCATCCTT	339
Exon20	ME 20f	TGGAGCTTTGGGAGGCTTTGT	
	ME_20r	GGGGCAGTGGGTCAATCAAC	344
Exon21	ME 21f	CCAAGAGGCTAAGGCTGGAG	
	ME_21r	AGGCTCAACGTTTTATGTTAATCAC	484
Exon22	ME 22f1	GCACAGCCAGAAGTACCAGTTTTGT	
	ME_22r2	TGGAGCAGCAGAGCTTTTATTATCT	382
Exon23	ME 23f	CCACTCGGCCAGACTTAGAAG	
	ME_23r	CCTGCCAGGATGGAAACTGTG	380
Exon24	ME 24f	TCGGGGAGAAAGGGGATGATT	
	ME_24r	GCAGACAGCAAACCTTAGGGAAA	347
Exon25	ME 25f	TGGCACCAGAACTATACCAAATATG	
	ME_25r	GATATGGTGCCAGTTGGGGTTT	486
Exon26	ME 26f	CCTCATGCAACGATCCTCCA	
	ME_26r	GGTATCCTGAAGTGCCTTGTCCA	448
Exon27	ME 27f	AAACACTCTGTTAAGTTGAAACAAC	
	ME_27r3	GTTCTGTGTTTTTCCCCCTGTGCTT	383
Exon28	ME 28f1	CATGGGAACCCATTTGTGTTATATT	
	ME_28r2	GGGAGAGCTAGCCTGTTGTTTCAA	641

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

Table 1b: PCR amplicon information for *ABCB11*

Fragment	Primer designation	5´ - 3´Sequence	Product size (bp)
Prom9	BS85	CCCCACTCCCAGAGTTTCTGATT	
	BS86	ACATCAGATGGTGTAAATGGTATGGAGA	500
Prom8	BS03	TGGCAACATCCTTCTCCCCTC	
	BS04	CATGCTGGCCAGACACATCTG	444
Prom7	BS05	GGGACCACAACCCTACCTATTTTCA	
	BS06	ACAGCAGTCAAAGGGTAAAAATGCC	388
Prom6	BS07	AAAACTCTATGCTTTTATCTTCCCTTGGAT	
	BS08	GTTCAAATCTCAAATATATCAAGGTCAGCAA	390
Prom5	BS09	TTGGTGCCAAATCCAATATTACTACATTT	
	BS10	ACATGCCAGTCTGTGCTAACCTTCT	374
Prom4	BS11	CAAATTAAGAAGCACTGGCCCATC	
	BS12	TGAGCTTACTTTTCCCTAAATGCCTCT	325
Prom3	BS91	TAGGGCTGTTGAGAGGCCCAAAT	
	BS92	CAGATGCTCCTGGCCCAGCTA	439
Prom2	BS15	ATACAGAGCTTCATCTGGTGTGTCCA	
	BS16	CATACGAGAAACATACTCAGAGCAATAACC	432
Prom1+Exon1	BS81	AAAATAAGGGTTGGGATAGCCTGAAT	
	BS82	GTAGCGTTGTTTACACTGGACAAGAGAC	487
Exon2	BS19	TGTCTGAGATTGATTAGTAAAGATTATTTCTTGAAC	
	BS20	TGCTGATTTTTTTCTGCTCCTTGAA	322
Exon3	BS77	GCAGGAAGAAAGAAAAGGCAAATAGTC	
	BS78	GCAATGTGCATGAAAGTATTCTCTGCT	275
Exon4	BS23	CTAAAGAATAACTCATAAAATCACCACCTAGGG	
	BS24	CACTGAAGTTAAAAACCTGCCAATATGACTA	252
Exon5	BS25	ATTCCCATGAAATTTGGTGTTGAGTACTA	
	BS26	AGCTCAGCCAGTAAAATCCCCTCTATA	449
Exon6	BS89	CTGGTGGCTTGATCCTACTTGCA	
	BS90	CTCCATTCTCTCATGTCCTCTGGAA	283
Exon7	BS29	TCTGAATTACTTTCCCCCTTTTCTCAA	
	BS30	GATGAAACATAGGAAAACTGAAAATATTTTTAAAT	361
Exon8	BS31	AGGGTGATAGGGATAGAGAGATGGGA	
	BS32	GCTAACTGTACTCAGGAAAAGGGACTCA	384
Exon9	BS33	TGAGAATCTAATATTGTATTAAACCCATGCC	
	BS34	CAAGGTGGGTCTGCCGCTT	333
Exon10	BS35	GCTGCTCTGTGTTTGCGATGATT	
	BS36	TCCTGAAGGCACCAAAGTAATAAACAA	396
Exon11	BS95	GATAACTCTCTGCGTTAACATGGAAGACC	
	BS96	AACTTTAAGGTAAATTCTTCAGGAGTTCATTCTG	301
Exon12	BS39	CTAAGAGGCACAATAAATGTATACATCTTCATTT	
	BS40	GAAACAGAGTCAGGCTTCAGAAAATGA	387
Exon13	BS41	CACACAGACACCGAGTATCAACACAA	
	BS42	CAGGACAGTCTCAATGTATGCTACACCT	333
Exon14	BS43	AGAATCTTATTGGCCTCTATTTTTTCTGC	
	BS44	TTGGGAATCATACGAGAAGAAATGTGTA	411
Exon15	BS45	CGGAGCAGTGAAAATTCGTGC	

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

	BS46	TCAACTACTCCCATCCCTCCCAC	392
Exon16	BS47	TATTCCCAGGGACCAAGGATCTAACTT	
	BS48	GTTGTTGGGAGAACAGTGAGTATTGAAA	416
Exon17	BS49	CATAGTAAAGAATTCTACTTGGATATGGTTCTGTT	
	BS50	CTCAGAGTTTCCTTGTTGTACCTGAGATATT	247
Exon18	BS51	TACCATGTGACCTACCAAACATTTCTATTTATT	
	BS52	CCTAGAAAACTCATATTCTCAGGCTTAAAGG	318
Exon19	BS99	AACTAATCAAACTATGACTCTTAAATCTGTGAATGC	
	BS100	GAGAAAAATGAACTAACAGAAAAGAGCTAAATACATG	382
Exon20	BS103	CTTACATTAGGGTTCACTCTTGGTGTTGTA	
	BS106	TTTGTTAGGAGCTCCCCAGATTTAGCAA	383
Exon21	BS57	TGTATCTCTTTTCACCCAATTTCTACAGGT	
	BS58	GGTCCCTATTCCATAGAAAACATGCA	381
Exon22	BS59	CATTCTGCCATTTGCACCATCTAAT	
	BS60	GCTTGGGATGGTTTGCTAAGCA	443
Exon23	BS61	ATATTTTATCACAACTTTATAAAAGGTCTGACCTTTT	
	BS62	GCTTCCTTCAGTCTCTTCGTACTACTCG	401
Exon24	BS63	ATATTTGGTCCTTTCCTGGCAGAAC	
	BS64	CCCACACCATCCCCTGACA	488
Exon25	BS65	TTTGGCAGCATGGTTTGAAGG	
	BS66	GAGTCTGGCAAAGCAAAAACTTGAAG	407
Exon26	BS67	GGGATTGTTAGTCTGTTAAGCAAACCAA	
	BS68	GCTCAACCTGTACACTCTGGTCATTCTAC	406
Exon27	BS107	ATTACAGAGGAGACCTTGACATGAGTTCAG	
	BS108	CGAATCAGAAAATTGAAAATAGTGCCATT	334
Exon28	BS75	CAAGTATAGGATTGTTATTCAGGTCGTGTTAA	
	BS76	TTGGGTTTTCCCTCACATGGAC	556

Table 2a: Polymorphisms identified in the *ABCB4* gene

Variant ID	5' Sequence	Genetic variation	3' Sequence	Region	Amino acid change	CA		KO		JA		Total	
						n	%	n	%	n	%	n	%
43	TCAATGCAC	g.-7676A>T	GTCTCACAA	PromA		110	0.9	96	0.0	88	0.0	294	0.3
44	CTACCCTCT	g.-7554T>C	CAATGCCTC	PromA		110	0.9	96	0.0	88	0.0	294	0.3
45	GAGTGAAGT	g.-7253G>A	TAGAAATCT	PromA		106	0.0	96	1.0	94	0.0	296	0.3
46	AATTTAGAA	g.-7114A>T	ACTCAATAG	PromA		108	0.0	96	1.0	94	0.0	298	0.3
1	AAGAGGAAA	g.-7094G>C	TTTCTTGTA	PromA		108	13.0	96	30.2	94	24.5	298	22.1
47	CAAGAATTT	g.-6816C>T	ATTAGGCAA	PromA		102	0.0	96	1.0	92	0.0	290	0.3
48	GAGAGAGAG	g.-6639A>C	GAGCTGAAT	PromA		110	0.9	90	0.0	92	0.0	292	0.3
49	GAGAGAGAG	g.-6637_-6636 delAG	CTGAATCAG	PromA		110	0.0	90	1.1	92	0.0	292	0.3
2	GTGCCTTTG	g.-6588G>T	GTGTGCTGG	PromA		110	2.7	88	0.0	92	2.2	290	1.7
3	AAAGAAGAA	g.-6540C>T	GAAACCAAA	PromA		108	14.8	86	26.7	90	27.8	284	22.5
4	TTAGTGACC	g.-6325A>G	AAAGTTTGG	PromA		108	17.6	90	31.1	92	26.1	290	24.5
5	ATTCTTTTT	g.-6014G>T	TACAAACCC	PromA		108	16.7	94	28.7	88	26.1	290	23.4
50	ACTGGTGCT	g.-5941G>A	TGGGCACTA	PromA		108	0.0	94	0.0	88	1.1	290	0.3
6	TGAAGTCAC	g.-5859G>A	TGGCCAGAG	PromA		108	16.7	94	28.7	88	26.1	290	23.4
7	ATGAGATGA	g.-5717T>C	ATATATGTG	PromA		110	0.0	92	1.1	86	3.5	288	1.4
8	CCTTCTTTA	g.-5610T>C	ATGCCTAAA	PromA		110	100.0	96	100.0	94	97.9	300	99.3
9	TAAGTGTGG	g.-5570G>C	CAGCAATTA	PromA		110	12.7	96	29.2	94	24.5	300	21.7
51	TACTCTCAC	g.-5509G>A	GCTCTTATG	PromA		110	0.0	96	1.0	94	0.0	300	0.3
10	CTCTCTTGT	g.-5236C>T	TGAGTAATA	PromA		108	15.7	90	27.8	94	26.6	292	22.9
52	GAGGATAAA	g.-2551A>T	AAGAAAGAT	PromB		126	0.0	88	0.0	90	1.1	304	0.3
11	AGCCTTACA	g.-2478T>G	CAATGCATA	PromB		126	4.8	88	0.0	90	2.2	304	2.6
12	GAAGGAATT	g.-1921T>C	GGGTTGATT	PromB/Exon -3		122	4.9	92	0.0	82	1.2	296	2.4
53	GAAGAGAAT	g.-1899C>A	CTCATGGTC	PromB/Exon -3		122	0.8	92	0.0	82	0.0	296	0.3
13	ATCCTAATA	g.-1603A>T	CACCCTTAT	PromB		128	0.0	94	2.1	86	1.2	308	1.0
14	TTTATAGAT	g.-1584C>T	CAATGACTG	PromB/Exon -2		118	11.0	94	29.8	74	23.0	286	20.3
54	ACACCAGGG	g.-1510T>G	CCACCCAGC	PromB		126	0.0	94	1.1	68	0.0	288	0.3
15	CTTATACCA	g.-1484T>C	GCTCTGCTT	PromB		126	0.0	94	1.1	68	5.9	288	1.7
55	TTTGAAAGT	g.-1146C>T	TCCGGTTTC	PromB		126	0.0	92	1.1	82	0.0	300	0.3

16	TGGTAGGAG	g.-1031C>T	AGAGACAAT	PromB		126	11.1	92	30.4	82	24.4	300	20.7
56	GAGACAATT	g.-1020C>G	AATACAGAC	PromB		126	0.0	92	1.1	82	0.0	300	0.3
17	ATTCAATAC	g.-1014A>G	GACAGAAGT	PromB		126	13.5	92	30.4	82	24.4	300	21.7
18	GAACTGGGG	g.-682A>C	TGCGGAAGC	PromB/Exon -1		124	1.6	70	0.0	64	0.0	258	0.8
19	AGGCTCCAG	g.-495C>G	CTGATCTCG	PromB		126	17.5	86	29.1	84	28.6	296	24.0
20	GCGCCCCGG	g.-414C>T	GGCAAGAGC	PromB		126	4.8	86	3.5	84	6.0	296	4.7
21	GGCAGGCTG	g.-395C>G	GCCCCTGGC	PromB		126	13.5	86	3.5	84	6.0	296	8.4
22	GCCCGCGCC	g.-378T>C	AGCCTGGGG	PromB		126	26.2	86	27.9	84	20.2	296	25.0
57	GCGTTTCCC	g.-292G>T	GGCCGGACG	PromB		128	0.0	96	0.0	92	1.1	316	0.3
58	CCGGACGCG	g.-280C>A	GTGGGGGGC	PromB		126	0.8	86	0.0	84	0.0	296	0.3
59	CCCTGCCAG	g.-186A>G	CACGCGCGA	PromB/Exon 1		128	0.0	96	1.0	92	0.0	316	0.3
23	TGCCCCCGG	g.-75G>T	CCCCGCGAC	PromB		128	0.0	96	0.0	92	1.1	316	0.3
24	TTTATGTCG	g.12597C>T	TGGGTACCA	Exon 4	L59L	126	12.7	96	25.0	94	21.3	316	19.0
60	CAAATTTGT	g.12680T>G	GATACTGCA	Exon 4	V86V	126	0.0	96	0.0	94	1.1	316	0.3
61	ATTTGTTGA	g.12683T>G	ACTGCAGGA	Exon 4	D87E	126	0.0	96	0.0	94	1.1	316	0.3
62	TTCTCCTTT	g.12705C>T	CAGGTAAGC	Exon 4	P95S	126	0.0	96	0.0	94	1.1	316	0.3
63	TAGCTTTCA	g.20782T>G	ACATTTAAA	Intron 4		114	0.0	88	1.1	70	0.0	272	0.4
25	TTTTAAAAA	g.20813C>T	CTGGCAATG	Intron 4		116	3.4	90	1.1	70	0.0	276	1.8
26	TCACCTATT	g.21044A>G	TTATCATTT	Intron 5		122	16.4	52	15.4	46	32.6	220	19.5
27	AAAAGAAAA	g.22281_22284delGAAAA	AAGAAAAGA	Intron 5		126	7.9	88	0.0	84	0.0	298	3.4
28	TGACATCAA	g.22490C>T	GACACCACT	Exon 6	N168N	126	42.9	88	37.5	90	44.4	304	47.7
29	GAACTCAAT	g.22509A>G	CGCGGCTAA	Exon 6	T175A	126	3.2	88	0.0	90	0.0	304	1.3
64	CTCTGCAGC	g.23831C>T	GTTTGGGCA	Exon 7	A232A	128	0.0	94	0.0	90	1.1	312	0.3
65	AAGGGTTGA	g.25313C>G	CAGAGTGCC	Intron 7		124	0.8	96	0.0	90	0.0	310	0.3
30	TGTCCAGAT	g.25376A>T	CTCTCGGCA	Exon 8	I237l	126	15.1	96	27.1	90	27.8	312	22.4
66	GTTAATATA	g.28354T>C	GCATCATAT	Exon 9	Y309Y	128	0.8	96	0.0	92	0.0	316	0.3
67	GCATATGTG	g.30584A>G	TCTTTGATA	Exon 10	I367V	118	0.8	92	0.0	92	0.0	302	0.3
68	TAATATTTA	g.31823T>G	AGGAATTCC	Intron 11		128	0.0	96	0.0	92	1.1	316	0.3
31	ACTTTTTTT	g.31941delT	CAAATTTCA	Intron 11		128	7.0	96	3.1	94	1.1	318	4.1
69	ACCCTGATG	g.32140A>G	GGGCACAGT	Exon 12	E450G	128	0.0	96	1.0	94	0.0	318	0.3
70	ACAAATTTG	g.32232C>T	GTGTGAATC	Intron 12		128	0.0	96	1.0	94	0.0	318	0.3
32	GGCAATGCC	g.32277G>T	ATGGATAAT	Intron 12		124	6.5	96	3.1	94	3.2	314	4.5

33	CAGCTATTA	g.35024A>G	ATGGTTAAA	Intron 12		124	95.2	94	70.2	94	72.3	312	80.8
71	ACAGGTAAA	g.35273G>A	CCTCTGATA	Intron 13		126	0.0	96	0.0	94	1.1	316	0.3
72	AGTGTGCCA	g.43862A>G	TACTGTAAC	Intron 14		122	0.8	88	0.0	72	0.0	282	0.4
34	TGTGCCAAT	g.43864A>G	CTGTAACCC	Intron 14		124	0.0	88	1.1	72	2.8	284	1.1
73	TAGCACACC	g.43938G>A	ACTGTCTAC	Exon 15	R590Q	124	0.8	88	0.0	74	0.0	286	0.3
35	GCTGCCACT	g.48606A>G	GAATGGCCC	Exon 16	R652G	124	7.3	86	2.3	74	1.4	284	4.2
36	GCTACAATT	g.48771A>G	TTGAAATTC	Intron 16		114	5.3	86	2.3	70	1.4	270	3.3
37	TTGCAAACA	g.51576C>T	CACATAACA	Intron17		122	95.1	70	74.3	80	71.3	272	82.7
38	TTACATAAC	g.56969A>G	TGGTTTTAG	Intron20		126	6.3	60	3.3	66	1.5	252	4.4
74	ATATTTTAC	g.58304T>C	GTATTAATG	Intron21		124	0.0	96	0.0	92	1.1	212	0.3
39	CTGTATTAA	g.58312T>C	GTCTAGAAC	Intron 21		122	4.1	96	1.0	92	1.1	310	2.3
75	AAAGGAGGC	g.63395delT	GAAGAGATG	Intron 22		124	0.8	92	0.0	94	0.0	310	0.3
76	AATATTAAG	g.63598A>T	TTATTCTAT	Intron23		124	0.0	92	1.1	94	0.0	310	0.3
40	ATGGTCAAG	g.68988A>G	AGCAAAGAA	Exon 26	E1099G	112	1.8	92	0.0	84	0.0	288	0.7
41	TATTTATAA	g.72169T>C	TTGGTTAAC	Intron 26		122	91.8	90	65.6	92	75.0	304	78.9
42	GTAACATTT	g.73092T>C	CAAATTTAC	Intron 27		124	7.3	94	1.1	86	1.2	304	3.6

Numbering of nuccleotide according to GenBank accession numbers AC005068.2 (non-coding exons –3 to 1 and coding exons 2 and 3), AC006154.1 (exon 4-12) and AC0005045.2 (exon 13-28). Positions of polymorphisms are counted with respect to the MDR3 translation start site (Exon 2) with the A of ATG being +1 and the 5'-following base being –1. n: number of alleles analyzed (number of subjects times 2). %: allele frequencies for Caucasians (CA), Japanese (JA) and Koreans (KO). The numbers 1-42 in the variant ID column indicate all variants included in haplotype analysis and linkage disequilibrium estimation.

Table 2b: Polymorphisms identified in the *ABCB11* gene

Variant ID	5' Sequence	Genetic variation	3' Sequence	Region	Amino acid change	CA		AA		JA		Total	
						n	%	n	%	n	%	n	%
54	GTAGTCACA	g.-15595C>T	TTTCAGAGC	Promoter		186	0.5	92	0.0	88	0.0	366	0.3
1	ACACTCTCT	g.-15281_-15278 delCTCT	CACACAGCA	Promoter		186	10.2	92	4.3	86	26.7	364	12.6
2	CCCCCTCCC	g.-15150T>C	GCCCCCAGA	Promoter		148	48.0	76	39.5	92	25.0	316	58.9
3	TGACTGTAG	g.-15018G>A	GACCACAAC	Promoter		158	10.1	78	0.0	92	29.3	328	13.1
4	ATTAAGCAC	g.-14944G>A	ATCAACTCA	Promoter		198	10.1	72	0.0	96	28.1	366	12.8
5	CTATTGGGA	g.-14589A>T	TCTTTTCCC	Promoter		198	0.0	90	2.2	88	0.0	376	0.5
55	TGAAGCAAA	g.-14524A>T	TTTTTTTCC	Promoter		198	0.0	90	1.1	88	0.0	376	0.3
6	TACATTTGC	g.-14473G>A	TCAACTCAG	Promoter		198	2.5	90	18.9	88	0.0	376	8.8
7	TTGCATAGA	g.-14437G>A	GAAACATCT	Promoter		198	29.8	94	22.3	96	14.6	388	24.2
8	ATTATATGT	g.-14353T>C	ATAATTTTG	Promoter		190	61.6	80	92.5	88	75.0	358	71.8
56	ATAAACCAT	g.-14316C>A	TTATACATA	Promoter		192	0.5	80	0.0	88	0.0	360	0.3
9	ACCATCTTA	g.-14312T>C	ACATAAATT	Promoter		192	0.0	80	0.0	88	3.4	360	0.8
57	ATAAATTCC	g.-14300A>T	ATAGAGAAA	Promoter		192	0.0	80	1.3	88	0.0	360	0.3
10	TTTAATTTC	g.-14207T>C	GCAAATTAA	Promoter		190	2.1	80	17.5	88	10.2	358	7.5
11	TTGTTACAC	g.-14104C>T	TTAGGAGGA	Promoter		196	2.6	92	0.0	96	0.0	384	1.3
12	CATGATAGC	g.-14035A>G	CCCAACTCC	Promoter		194	1.5	92	1.1	96	0.0	382	1.0
58	AAGGCTGGA	g.-13910G>A	TGAGAGGCA	Promoter		202	0.0	94	1.1	96	0.0	392	0.3
13	AGAGGAAGA	g.-13814G>A	GCAGCACAA	Promoter		194	0.0	94	6.4	88	0.0	376	1.6
14	GCACAAATA	g.-13801T>C	ATTGGAGCT	Promoter		194	1.5	94	0.0	88	0.0	376	0.8
15	CTCAGACTT	g.-13662T>C	TGAGCAAGG	Promoter		192	0.0	94	7.4	86	0.0	372	1.9
83	TTAAAGGTA	g.- -13523[T]9	GTCTTGTTA	Promoter		200	10.0	90	11.1	96	28.1	386	14.8
84	TTAAAGGTA	g.- -13523[T]10	GTCTTGTTA	Promoter		200	9.0	90	18.9	96	6.3	386	10.6
85	TTAAAGGTA	g.- -13523[T]11	GTCTTGTTA	Promoter		200	65.0	90	53.3	96	45.8	386	57.5
86	TTAAAGGTA	g.- -13523[T]12	GTCTTGTTA	Promoter		200	16.0	90	16.7	96	19.8	386	17.1
59	CTGGGCCAG	g.-13595G>A	AGCATCTGG	Promoter		198	0.0	94	1.1	96	0.0	388	0.3
16	CAAGCACAC	g.-13333T>C	CTGTGTTTG	Promoter		196	0.0	76	0.0	96	3.9	368	0.9
17	ATGTTTCTC	g.-13297G>A	TATGTCACT	Promoter		196	0.0	76	3.9	96	0.0	368	0.8
60	TCCACAGTG	g.-13142G>A	AGTCCATTA	Exon 1		194	0.0	76	0.0	92	1.1	362	0.3

18	TTGATTAAA	g.-77G>A	AAGAAAGAA	Intron 1		202	2.5	88	11.4	90	12.2	380	6.8
19	ATTTTTTTT	g.1319delT	CTGACAGAT	Intron 2		198	0.0	88	3.4	92	0.0	378	0.8
20	TTTAAATCC	g.3754T>C	TATGTTTTT	Intron 3		198	7.1	62	32.3	88	12.5	348	12.9
21	GTTACAAGA	g.3781T>C	GAGAAGAAA	Exon 4	D36D	198	0.0	62	0.0	88	26.1	348	6.6
22	GAATCTAGT	g.4542A>T	ACTAAATTA	Intron 4		184	0.0	90	0.0	92	2.2	366	0.5
61	CAAGTTTCG	g.4621G>A	TTTTCTTCA	Exon 5	R52R	184	0.0	90	1.1	92	0.0	366	0.3
23	AGATGTTTT	g.4735T>C	ATTGACTAC	Exon 5	F90F	184	2.7	90	13.3	92	12.0	366	7.7
24	GGGTAGGTT	g.4862G>A	TTTTTGTTT	Intron 5		182	4.9	90	2.2	94	0.0	366	3.0
25	GCTGAACAT	g.21416C>T	GAGAGCGAA	Exon 6	I134I	192	0.0	86	18.6	88	0.0	366	4.4
26	AGCTCCTCC	g.21507G>A	TATAATTTA	Intron 6		196	0.0	92	18.5	92	0.0	380	4.5
27	ACAATGAGA	g.21554T>G	GCAATGTGT	Intron 6		196	0.0	92	0.0	92	4.3	380	1.1
62	TGTATTGAA	g.22567A>T	GTACTTTCT	Intron 6		198	0.5	94	0.0	94	0.0	386	0.3
63	TTTGAATGA	g.24203T>C	CAAATTCAG	Intron 7		192	0.5	90	0.0	94	0.0	376	0.3
64	TCTAGTGAT	g.24248A>G	TTAATAAAA	Exon 8	I206V	194	0.0	90	1.1	96	0.0	380	0.3
28	TACGGACTA	g.27224T>C	GAGCTGAAG	Exon 9	Y269Y	200	0.0	80	0.0	96	27.1	376	6.9
65	CTGATGAAG	g.27268T>C	CATTTCATC	Exon 9	V284A	200	0.5	80	0.0	96	0.0	376	0.3
66	GTGAGAAAA	g.27313G>A	AGAGGTTGA	Exon 9	R299K	200	0.0	80	0.0	96	1.0	376	0.3
29	ACTGCATCA	g.31773C>T	GGCCTGTTT	Intron 9		178	5.1	70	1.4	48	0.0	296	3.4
30	TGTTTCTGC	g.31811C>T	GAAATTGAC	Intron 9		196	4.6	72	4.2	86	0.0	354	3.4
31	TTGACTCAA	g.31825G>A	CATTTTGTC	Intron 9		196	4.6	72	26.4	86	0.0	354	7.9
32	GACTCAAGC	g.31827A>G	TTTTGTCTT	Intron 9		196	70.4	72	84.7	86	90.7	354	78.2
33	TAGAAAAGG	g.31890A>G	ATAGTGATG	Exon 10	G319G	196	4.6	72	26.4	86	0.0	354	7.9
34	GACTTATTG	g.32034A>T	CCGAGACAT	Intron 10		196	0.0	66	4.5	82	0.0	344	0.9
67	CCTCAGTGT	g.38161C>T	ATAGTAGGA	Exon 11	V366V	196	0.0	88	1.1	94	0.0	378	0.3
35	CATTTTTGA	g.38248G>A	ACAATAGAC	Exon 11	E395E	196	0.0	88	8.0	96	0.0	380	1.8
36	GCAGAGATA	g.41348C>T	GCCAAAGAT	Intron 11		198	0.0	72	2.8	80	0.0	350	0.6
37	CCACAAATT	g.41622G>T	CTCATTTTC	Intron 12		196	1.0	72	0.0	74	0.0	342	0.6
38	CAGTGACAA	g.44255delT	CTGAACTTT	Intron 12		190	0.0	94	2.1	92	0.0	376	0.5
39	TCAACATGG	g.44308T>C	CATTAAACC	Exon 13	V444A	190	59.5	90	65.6	92	80.4	372	66.1
40	TTGATCAAA	g.44481C>T	AGAAAGGTG	Intron 13		188	59.0	90	65.6	92	80.4	370	65.9
68	CAAGGAGGC	g.46246C>T	AATGCCTAC	Exon 14	A535A	184	0.0	86	0.0	96	1.0	366	0.3
41	GGGAGAAAC	g.46311T>C	AAGAGGTCG	Intron 14		182	60.4	86	66.3	94	79.8	362	66.9

42	GTTGCTCAT	g.48611C>G	GCTTGTCTA	Exon 16	R616G	194	0.0	90	2.2	96	0.0	380	0.5
69	CGCTTGTCT	g.48620A>G	CGGTCAGAG	Exon 16	T619A	194	0.0	90	1.1	96	0.0	380	0.3
70	CAGAGCTGC	g.48634A>G	GATACCATC	Exon 16	A623A	194	0.0	90	1.1	96	0.0	380	0.3
43	GAAGATGAC	g.49653A>G	TGCTTGCGA	Exon 17	M677V	190	4.2	86	14.0	88	0.0	364	5.5
71	CCGGCAAC	g.53835G>A	CTCCAAGTC	Exon 18	R698H	196	0.5	82	0.0	94	0.0	372	0.3
72	GAACCTCCA	g.53876T>C	TAGCTGTTG	Exon 18	L712L	196	0.5	82	0.0	94	0.0	372	0.3
44	TTAATATAA	g.59981C>A	CCTCTCTCT	Intron 18		192	43.2	84	21.4	90	27.8	366	34.4
45	AATAGATTT	g.73116_73119delATTT	TTCTATTTA	Intron 19		192	0.0	66	6.1	96	0.0	354	1.1
46	ATTTATAAT	g.73132_73133insCAA	AAAGTTACT	Intron 19		194	0.0	66	6.1	96	0.0	356	1.1
47	ACTTTCTTG	g.73148T>C	TTACTATCT	Intron 19		196	69.4	66	93.9	96	0.0	358	82.1
48	TTTTGTAGC	g.73233A>G	ATGGGCTGT	Exon 20	A804A	200	0.0	66	93.9	96	0.0	362	1.1
49	CTACAGATG	g.73505C>T	TTCCCAAGT	Exon 21	A865V	178	0.0	74	0.0	82	2.4	334	0.6
73	TGGGAGGGG	g.73539A>C	AATAGAAGT	Intron 21		176	0.6	74	0.0	84	0.0	334	0.3
74	TGGTAAAAG	g.81650C>T	GACTGTGTG	Intron 21		196	0.0	86	1.2	94	0.0	376	0.3
75	GTCACGAAA	g.82653G>A	GAGTTATTT	Intron 22		136	0.0	90	0.0	46	2.2	272	0.4
50	GTTATTTCT	g.82665G>A	CCCTTGTAT	Intron 22		202	0.0	92	6.5	82	0.0	376	1.6
76	AGGAGAGGC	g.82759G>A	GTTCATTGA	Exon 23	R958Q	202	0.0	92	0.0	94	1.1	388	0.3
77	CAATATTTA	g.82829C>T	GGATTCTGC	Exon 23	Y981Y	200	0.0	92	1.1	94	0.0	386	0.3
78	GTTTATTGC	g.82871G>A	AATTCTGCT	Exon 23	A995A	200	0.0	92	1.1	94	0.0	386	0.3
51	ACTGAGTGC	g.85620A>G	ACAGCTCTT	Exon 24	A1028A	194	54.1	76	26.3	90	47.8	360	47.8
79	CTATGCAGC	g.85774C>A	ATAAAAAAG	Intron 24		194	0.0	76	0.0	90	1.1	360	0.3
80	GCAAACTAA	g.85848C>G	ACAAGAACA	Intron 24		194	0.0	76	1.3	90	0.0	360	0.3
52	TAAATTTAC	g.87308A>G	TATCCTTCT	Exon 25	T1086T	186	0.0	76	7.9	82	0.0	344	1.7
53	CAATCATGC	g.94270A>G	TCTTTGCAT	Intron 27		172	55.2	64	75.0	94	51.1	330	48.2
81	CCAGAACGC	g.94387G>A	GATATCATT	Exon 28	A1283A	194	0.0	64	0.0	94	1.1	352	0.3
82	TCCCAGCAG	g.94582G>A	AGGGATTGT	Exon 28		192	0.0	26	0.0	88	1.1	306	0.3

Numbering of nucleotide according to the GenBank accession numbers AC008177.3 (promoter and exon 1-21) and AC069165.2 (newer classification AC069137.6)(exon 22-28). Positions of polymorphisms are counted with respect to the BSEP translation start (Exon 2) with the A of the start ATG being +1 and the 5'-following base being –1. n: number of alleles analyzed (number of subjects times 2). %: allele frequencies for Caucasians (CA), African American (AA) and Japanese (JA). The numbers 1-53 in the variant ID column indicate all variants included in haplotype analysis and linkage disequilibrium estimation.

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

Table 3: Prediction of functional consequences of non-synonymous SNPs in *ABCB4* and *ABCB11*

Amino acid change	Scoring systems for non-synonymous variants				
	Grantham	**SIFT**	**PolyPhen**	**Blosum62**	**EC/EU**
MDR3					
D87E	45	1.00	0.48	2	EC
P95S	74	0.48	0.87	-1	EC
T175A	58	0.01	0.72	-1	EC
I367V	29	0.23	0.96	3	EC
E450G	98	0.01	0.13	-2	EC
R590Q	43	0.01	2.51	1	EC
R652G	125	0.36	1.47	-2	EU
E1099G	98	0.04	1.58	-2	EC
BSEP					
I206V	29	1.00	0.23	3	EU
V284A	64	0.13	0.43	-2	EC
R299K	26	1.00	0.38	2	EU
V444A	64	0.63	0.78	-2	EC
R616G	125	0.01	3.16	-2	EC
T619A	58	0.00	1.78	-1	EC
M677V	21	0.29	0.82	1	EU
R698H	29	0.30	0.57	0	EC
A865V	64	0.02	1.12	0	EC
R958Q	43	0.04	0.24	1	EU

Grantham values range from 5 to 215; low values (<50) indicate chemical similarity and high values (≥50) indicate more radical differences. SIFT (sorting intolerant from tolerant) values range from 0 to 1; values close to 0 are less tolerated, while those near 1 are better tolerated substitutions. PolyPhen calculates PSIC (Position-Specific Independent Counts) scores for two amino acid variants in the polymorphic position. Large differences (>1.5) are less tolerated. BLOSUM62 values range from -4 to +3, with negative values indicating less acceptable and non-negative values indicating more acceptable substitutions. EC/EU classifies non-synonymous variants as evolutionary conserved (EC) or unconserved (EU) based on sequence alignments with mammalian orthologs.

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

Table 4: Summary statistics of population variation in the *ABCB4 and ABCB11* genes

Gen	Pop-ulation	SNP type	bp screened	n	variable sites*	Θ(x10^{-4}±SD)	π(x10^{-4}±SD)	*D*
ABCB4	Ca	All	16797	128	40	4.39±1.21	2.79±1.52	-1.11
		non-coding	12960	128	31	4.41±1.27	2.69±1.53	-1.16
		coding	3837	128	9	4.32±1.72	3.14±2.24	-0.66
		intron	7434	128	14	3.47±1.20	1.94±1.32	-1.17
		5´UTR	5526	128	17	5.67±1.87	3.70±2.31	-0.95
	Ja	All	16797	94	40	4.66±1.34	4.20±2.20	-0.31
		non-coding	12960	94	34	5.13±1.51	4.83±2.56	-0.18
		coding	3837	94	6	3.06±1.42	2.07±1.68	-0.73
		intron	7434	94	11	2.89±1.10	2.76±1.73	-0.12
		5´UTR	5526	94	23	8.98±2.57	7.62±4.23	-0.19
	Ko	All	16797	96	43	4.98±1.41	4.31±2.26	-0.43
		non-coding	12960	96	38	5.71±1.65	4.64±2.47	-0.59
		coding	3837	96	5	2.54±1.26	3.22±2.28	0.57
		intron	7434	96	12	3.14±1.16	1.49±1.09	-1.41
		5´UTR	5526	96	26	9.16±2.82	8.88±4.83	-0.09
ABCB11	Ca	All	12759	202	34	4.53±1.23	4.19±2.25	-0.21
		non-coding	8796	202	26	5.02±1.44	4.81±2.65	-0.12
		coding	3963	202	8	3.43±1.39	2.82±2.04	-0.39
		intron	6078	202	15	4.19±1.38	4.72±2.76	0.32
		5´UTR	2718	202	11	6.88±2.50	5.01±3.44	-0.65
	AA	All	12759	94	36	5.52±1.61	3.63±1.61	-1.07
		non-coding	8796	94	23	5.11±1.61	3.67±2.11	-0.84
		coding	3963	94	13	6.41±2.32	3.55±2.43	-1.22
		intron	6078	94	10	3.22±1.26	2.99±1.26	-0.18
		5´UTR	2718	96	13	9.35±3.39	3.39±3.55	-1.21
	JA	All	12759	96	31	4.73±1.41	5.28±2.78	0.36
		non-coding	8796	96	21	4.65±1.49	5.64±1.49	0.62
		coding	3963	96	10	4.91±1.92	4.48±2.90	-0.23
		intron	6078	96	11	3.52±1.34	3.75±2.30	0.17
		5´UTR	2718	96	10	7.16±2.80	9.87±5.87	0.97

CA, Caucasian; JA, Japanese; KO, Korean; AA, African American; * number of variable sites considered for population genetic calculations. The neutral parameter (θ), nucleotide diversity (π) and Tajima'*D* statistic were calculated as described by Tajima (Tajima 1989). SD = standard deviation

Table 5a: *ABCB4* haplotypes and estimated haplotype frequencies. Haplotypes are aligned according to sequence similarity. Variation and the reference haplotype are given at the top with colóur coding for promoter (blue), intronic (yellow), synonymous (green), and non-synonymous (red) polymorphic sites. Haplotypes with a frequency of ≥3% in at least one population are highlighted in bold letters. The frequencies of haplotypes that could be deduced unambiguouly from genotypes (no missing data, not more than one heterozygous variant) are marked with an asterisk. R, D and I indicate reference nucleotide, deletion and insertion.

	-7094G>C	-6588G>T	-6540C>T	-6325A>G	-6014G>T	-5859G>A	-5717T>C	-5610T>C	-5570G>C	-5236C>T	-2478T>G	-1921T>C	-1603A>T	-1584C>T	-1484T>C	-1031C>T	-1014A>G	-682A>C	-495C>G	-414C>T	-395C>G	-378T>C	-75G>T	12597C>T	20813C>T	21044A>G	22281_22284delGAAAA	22490C>T	22509A>G (T175A)	25376A>T	31941delT	32277G>T	35024A>G	43864A>G	48606A>G (R625G)	48771A>G	51576C>T	56969A>G	58312T>C	68988A>G (E1099G)	72169T>C	73092T>C	Frequency (%)		
Haplotype ID	**G**	**G**	**C**	**A**	**G**	**G**	**T**	**T**	**G**	**C**	**T**	**T**	**A**	**C**	**T**	**C**	**A**	**A**	**C**	**C**	**C**	**T**	**G**	**C**	**C**	**A**	**R**	**C**	**A**	**A**	**R**	**G**	**A**	**A**	**A**	**A**	**C**	**A**	**T**	**A**	**T**	**T**	**CA n=128**	**JA n=94**	**KO n=96**
MDR3_15								C																																			0.78		
MDR3_40								**C**														**C**																						**3.19**	**6.25**
MDR3_39								C														C																G							1.04
MDR3_13								C																				T															0.78	1.06	
MDR3_35								C																				T									T								1.04
MDR3_50	C							C																				T																	1.04
MDR3_30								C					T															T					G				T				C			1.06	2.08
MDR3_5								C								T						C											G				T				C		0.78		
MDR3_6								C									G											T					G				T				C		0.78		
MDR3_7*								C											G		G							T					G				T				C		0.78		
MDR3_8*								C											G		G												G				T				C		0.78		
MDR3_17								C														C						T					G				T				C		1.56		1.04

MDR3_18								C														C											G				T			G	C		1.56		
MDR3_19*								C														C											G				T						0.78		2.08
MDR3_20*								**C**														**C**											**G**				**T**				**C**		**17.97**	**11.70**	**18.75**
MDR3_9								C																		T			T					G				T				C		3.13		
MDR3_10								C																					T	G				G				T				C		3.13		
MDR3_11*								**C**																					**T**					**G**				**T**						**4.69**	**3.19**	**5.21**
MDR3_12*								**C**																					**T**					**G**				**T**				**C**		**32.81**	**36.17**	**25.00**
MDR3_14								C																										G				T				C		0.78	3.19	
MDR3_16*								**C**																				**D**	**T**					**G**				**T**				**C**		**7.81**		
MDR3_65							C	C																							T			G				T				C			1.06	
MDR3_66							C	C														C											G				T				C			2.13	
MDR3_42							C	C																					T					G				T				C				1.04
MDR3_4				G				C																					T					G				T						0.78		
MDR3_41								C								C													T					G				T				C			1.06	1.04
MDR3_58								C																					T					G								C			1.06	
MDR3_64								C								C						C											G				T				C			3.19	
MDR3_32								C																					T					G	G			T				C				1.04
MDR3_52								C												G														G	G			T				C			1.06	
MDR3_55								C													T	G									T														1.06	
MDR3_31								C													T	G									T			G				T				C				1.04
MDR3_53								C													T	G					G				T		T	G		G	G	T	G	C		C			1.06	
MDR3_54								C													T	G					G				T		T	G		G	G	T	G	C		C	C		1.06	
MDR3_21								C														C										D	T	G		G	G	T	G			C	C	1.56		
MDR3_22								C														C										D	T	G		G	G	T	G	C		C	C	1.56		
MDR3_23	**C**		**T**	**G**	**T**	**A**		**C**	**C**	**T**				**T**		**T**	**G**		**G**		**G**			**T**		**G**				**T**			**G**				**T**				**C**		**6.25**		
MDR3_24	C		T	G	T	A		C	C	T				T		T	G		G		G			T		G							G				T				C		0.78		
MDR3_25	C		T	G	T	A		C	C	T				T		T	G		G					T		G				T			G				T				C		0.78		3.13
MDR3_26	**C**		**T**	**G**	**T**	**A**		**C**	**C**	**T**				**T**		**T**	**G**		**G**					**T**		**G**				**T**											**C**		**2.34**	**6.38**	**1.04**
MDR3_68	C		T	G	T	A		C	C	T				T		T	G		G				T	T		G				T											C				

MDR3_69	C		T	G	T	A		C	C	T				T		T	G		G					T		G				T			G				T								
MDR3_43	C		T	G	T	A		C	C	T				T		T	G		G					T	T								G				T				C				1.04
MDR3_45	**C**		**T**	**G**	**T**	**A**		**C**	**C**	**T**				**T**		**T**	**G**		**G**					**T**		**G**				**T**														**10.64**	**15.63**
MDR3_47	C		T	G	T	A		C	C	T				T		T	G		G														G				T				C			3.19	4.17
MDR3_48	C		T	G	T	A		C	C	T				T		T	G					C		T		G				T															1.04
MDR3_27	C		T	G	T	A		C	C	T				T		T	G							T		G				T											C		0.78		
MDR3_28	C		T	G	T	A		C	C	T				T			G																G				T				C		0.78		
MDR3_29	C		T	G	T	A		C	C	T				T			G					C											G				T				C		0.78		
MDR3_49	C		T	G	T	A		C	C					T		T	G		G					T		G				T											C			1.06	
MDR3_67	C		T	G	T	A			C	T				T		T	G		G					T		G				T														1.06	
MDR3_1		T	T	G	T	A		C		T	G	C							G	T	G			T		G				T	D	T	G		G	G	T	G	C		C	C	1.56		
MDR3_2		T	T	G	T	A		C		T	G	C							G	T	G					G				T	D	T	G		G	G	T	G	C		C	C	0.78		
MDR3_3			T	G	T	A		C		T	G	C							G	T	G					G				T	D	T	G		G	G	T	G	C		C	C	0.78		

variable site promoter

variable site intron

variable site exon-synonymous

variable site exon-non-synonymous

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

Table 5b *ABCB11* haplotypes and estimated haplotype frequencies. Further details are as given for table 5a.

	-15281_-15278delCTCT	-15150T>C	-15018G>A	-14944G>A	-14589A>T	-14473G>A	-14437G>A	-14353T>C	-14312T>C	-14207T>C	-14104C>T	-14035A>G	-13814G>A	-13801T>C	-13662T>C	-13333T>C	-13297G>A	-77G>A	1319delT	3754T>C	3781T>C	4621G>A	4735T>C	4826G>A	21416C>T	21507G>A	21554T>G	27224T>C	31773C>T	31811C>T	31825G>A	31827A>G	31890A>G	32034A>T	38248G>A	41348C>T	41622G>T	44255delT	44308T>C (V444A)	44481C>T	46311T>C	48611C>G	49653A>G (M677V)	59981C>A	73116_73119delATTT	73132_73133insCAA	73148T>C	73233A>G	73505C>T (A865V)	82665G>A	85620A>G	87308A>G	94270A>G	Frequency (%)		
Haplotype ID	**R**	**T**	**G**	**G**	**A**	**G**	**G**	**T**	**T**	**T**	**C**	**A**	**G**	**T**	**T**	**T**	**G**	**G**	**R**	**T**	**T**	**A**	**T**	**G**	**C**	**G**	**T**	**T**	**C**	**C**	**G**	**A**	**A**	**A**	**G**	**C**	**G**	**R**	**T**	**C**	**T**	**C**	**A**	**C**	**R**	**R**	**T**	**A**	**C**	**G**	**A**	**A**	**A**	**CA n=202**	**AA n=94**	**JA n=96**
BSEP_1																																G												A			C				G		G	3.47		
BSEP_2																																G							C	T	C					C					G	0.99	1.06	
BSEP_3*																																**G**							**C**	**T**	**C**											**14.85**	**3.19**	
BSEP_4																																**G**							**C**	**T**	**C**					**C**				**G**		**G**	**1.98**		**8.33**
BSEP_5																																**G**							**C**	**T**	**C**					**C**							**2.97**		**6.25**
BSEP_6*																																**G**							**C**	**T**	**C**			**A**		**C**				**G**		**G**	**15.34**		**2.08**
BSEP_83																											C					G							C	T	C					C				G		G			1.04
BSEP_84																											C					G							C	T	C					C									1.04
BSEP_85																					C						C					G							C	T	C					C									3.13
BSEP_86																					C						C					G							C	T	C			A		C				G		G			2.08
BSEP_9								C																								G																						1.98		
BSEP_7								C				G																				G							C	T	C													0.99		
BSEP_8								C																								G					T		C	T	C					C									0.50		
BSEP_10								C																								G							C	T	C					C				G		G	0.99	1.06	
BSEP_11								**C**																								**G**							**C**	**T**	**C**					**C**							**2.97**	**11.70**	**1.04**
BSEP_12*								C																								G							C	T	C			A		C				G		G	3.96	4.26	
BSEP_13							A	C																								G					T		C	T	C					C							0.50		
BSEP_14		C						C																								G							C	T	C			A									0.50		
BSEP_15		C						C																								G												A			C				G		G	0.99		

BSEP_39					T			C																								G							C	T	C						C			A					1.06	
BSEP_42								C																								G		T													C								4.26	
BSEP_43								C																								G							C	T	C	G					C				G		G		1.06	
BSEP_44								C																								G							C	T	C		G		D	I	C	G		A					1.06	
BSEP_46								C																								G							C	T	C						C					G			3.19	
BSEP_48								C																								G							C	T	C						C			A					2.18	
BSEP_50								C																								G							C	T	C			A			C								2.18	
BSEP_51								C																																				A			C								1.06	
BSEP_88		C						C																			G					G							C	T	C						C									2.08
BSEP_89		C						C								C												C				G							C	T	C						C				G		G			1.04
BSEP_90		C						C								C					C	T						C				G							C	T	C						C									2.08
BSEP_91		C						C								C					C							C				G															C									2.08
BSEP_92		C						C								C					C							C				G							C	T	C						C				G		G			5.21
BSEP_93		C						C								C					C							C				G							C	T	C						C									2.08
BSEP_94		C						C								C					C							C				G							C	T	C			A			C				G		G			5.21
BSEP_16		C						C												C				A					T	T	A	G	G										G											1.98		
BSEP_17		C						C												C				A					T	T	A	G	G										G				C				G		G	0.99		
BSEP_18		C						C												C				A					T	T	A	G	G										G	A			C				G		G	0.99		
BSEP_19		C						C												C				A					T	T	A	G	G																					0.50		
BSEP_52								C												C					T	A					A	G	G														C								1.06	
BSEP_53		C			T		A	C																							A	G	G						C	T	C						C								1.06	
BSEP_54		C						C												C					T	A					A	G	G	T									G	A			C				G	G	G		1.06	
BSEP_55		C						C												C				A						T	A	G	G										G				C								1.06	
BSEP_56		C						C							C				D	C					T	A					A	G	G	T									G	A			C				G		G		1.06	
BSEP_57		C						C					A		C					C					T	A					A	G	G		A				C	T	C						C				G				1.06	
BSEP_58		C						C					A		C					C					T	A					A	G	G		A				C	T	C						C								3.19	
BSEP_59		C						C					A		C					C					T	A					A	G	G		A			D	C	T	C						C								2.18	
BSEP_20		C					A	C			T																																				C				G		G	1.49		
BSEP_21		C					A	C			T																																				C							0.50		

BSEP_22		C					A	C			T																																		A			C				G		G	0.50		
BSEP_23		C					A	C																								G												A			C				G		G	0.99		
BSEP_24		C					A	C																								G							C	T	C													0.50		
BSEP_25		C					A	C																																														4.46		
BSEP_26		**C**					**A**	**C**																																						**C**				**G**		**G**	**4.95**		**8.33**
BSEP_27		C					A	C																																						C							3.96		
BSEP_28		**C**					**A**	**C**																																			**A**			**C**				**G**		**G**	**8.91**		
BSEP_29		C					A	C																																			A			C						G	0.99		
BSEP_30		C					A	C																																			A			C							1.49		
BSEP_31		C					A	C						C																																							0.99		
BSEP_60		C					A																															C	T	C			A			C				G		G		1.06	
BSEP_61		C					A	C																								G							C	T	C						C								1.06	
BSEP_62		C					A	C																																						C								1.06	
BSEP_63		C					A	C																							A	G	G														C								5.32	
BSEP_64		C					A	C												C					T						A	G	G						C	T	C			A			C				G		G		1.06	
BSEP_65		**C**					**A**	**C**									**A**																														**C**								**8.51**	
BSEP_95		C					A	C																								G												A			C		T		G		G			1.04
BSEP_96		C					A	C																								G							C	T	C			A			C				G		G			3.13
BSEP_98		C					A	C																			G												C	T	C						C									1.04
BSEP_32		C				A		C		C								A		C			C									G							C	T	C													1.98	2.18	
BSEP_33		C				A		C		C								A		C			C									G							C	T	C						C				G		G	0.50	2.18	
BSEP_70		C				A		C		C								A		C			C									G							C	T	C	G		A			C				G		G		1.06	
BSEP_71		C				A		C		C								A		C			C									G							C	T	C		G				C								2.18	
BSEP_74		C				A		C		C								A		C			C									G							C	T	C						C								1.06	1.04
BSEP_75		C				A		C		C								A		C			C									G							C	T	C				D	I	C	G		A					2.18	
BSEP_76		C				A		C		C								A		C			C									G							C	T	C			A			C				G		G		1.06	
BSEP_99		C				A		C		C								A		C			C									G															C									5.21
BSEP_100		C				A		C		C								A		C			C									G												A			C				G		G			1.04
BSEP_101		C				A		C		C								A		C			C									G												A			C									1.04

DMD Fast Forward. Published on June 8, 2006 as DOI: 10.1124/dmd.105.008854
This article has not been copyedited and formatted. The final version may differ from this version.

BSEP_66		C				A		C																							A	G	G														C								1.06	
BSEP_67		C				A		C		C		G								C					T	A					A	G	G		A				C	T	C						C				G		G		1.06	
BSEP_68		C				A		C		C										C					T	A					A	G	G	T					C	T	C						C						G		1.06	
BSEP_69		C				A		C		C										C					T	A					A	G	G						C	T	C			A			C				G		G		2.18	
BSEP_34		C		A				C																																				A			C				G		G	1.98		
BSEP_77	D	C						C																								G							C	T	C		G		D	I	C	G							1.06	
BSEP_78	D	C						C											D	C					T	A					A	G	G														C								1.06	
BSEP_79	D	C						C		C										C					T	A					A	G	G														C								1.06	
BSEP_35	D	C	A	A				C																								G							C	T	C													1.98		
BSEP_36	D	C	A	A				C																								G							C	T	C						C				G		G	1.49		
BSEP_37	**D**	**C**	**A**	**A**				**C**																								**G**							**C**	**T**	**C**						**C**							**0.50**		**15.63**
BSEP_38	**D**	**C**	**A**	**A**				**C**																								**G**							**C**	**T**	**C**			**A**			**C**				**G**		**G**	**5.45**		**4.17**
BSEP_105	D	C	A	A				C																								G							C	T	C						C				G					1.04
BSEP_108	D	C	A	A				C																			G					G							C	T	C						C									2.08
BSEP_109	D	C	A	A				C	C																							G							C	T	C			A			C				G		G			3.13
BSEP_103		C	A	A				C								C					C							C				G							C	T	C						C									1.04
BSEP_104		C	A	A				C								C					C							C				G							C	T	C			A			C				G		G			1.04

variable site promoter
variable site intron
variable site exon-synonymous
variable site exon-non-synonymous

Fig. 1a

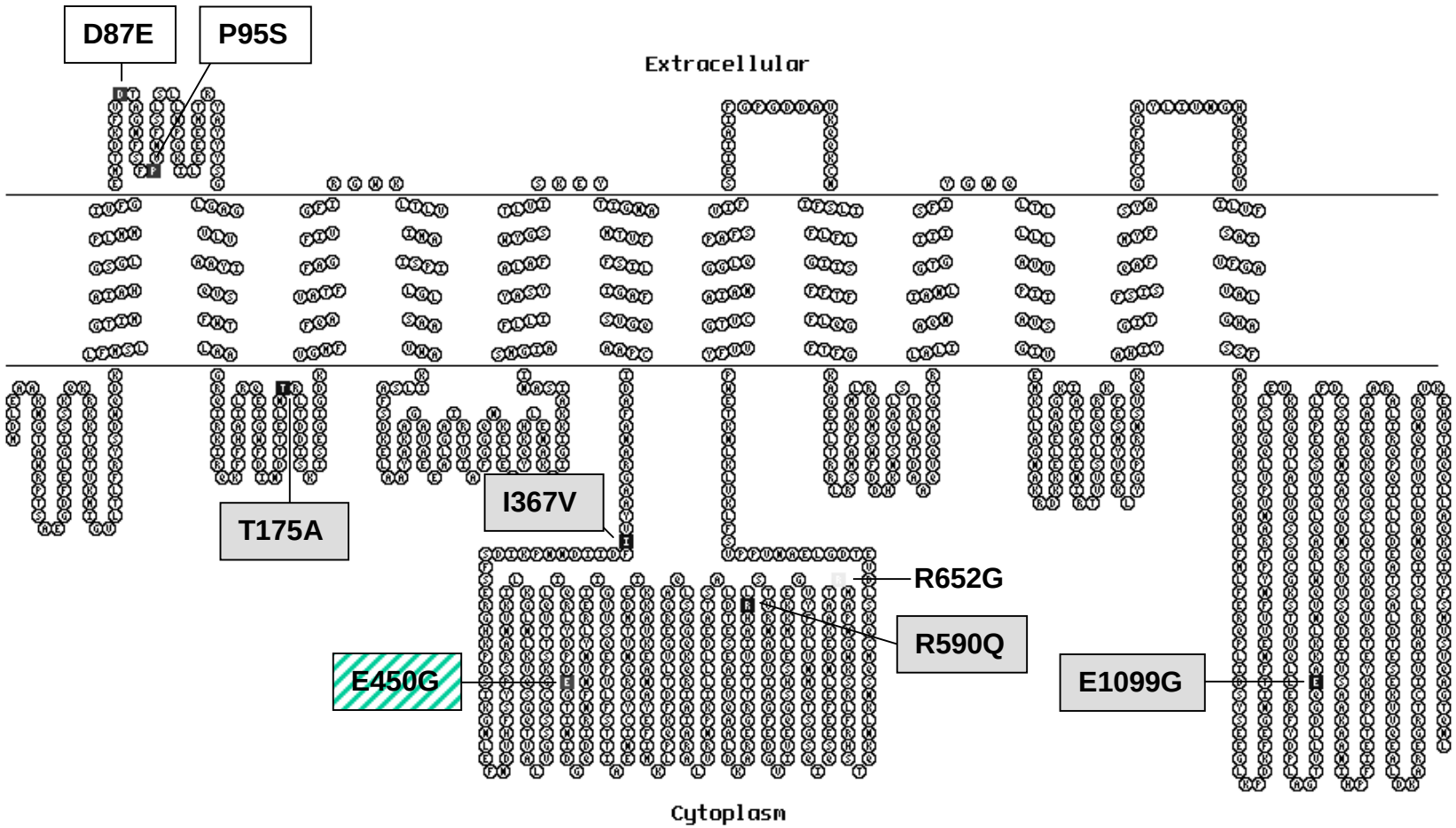

Fig. 1b

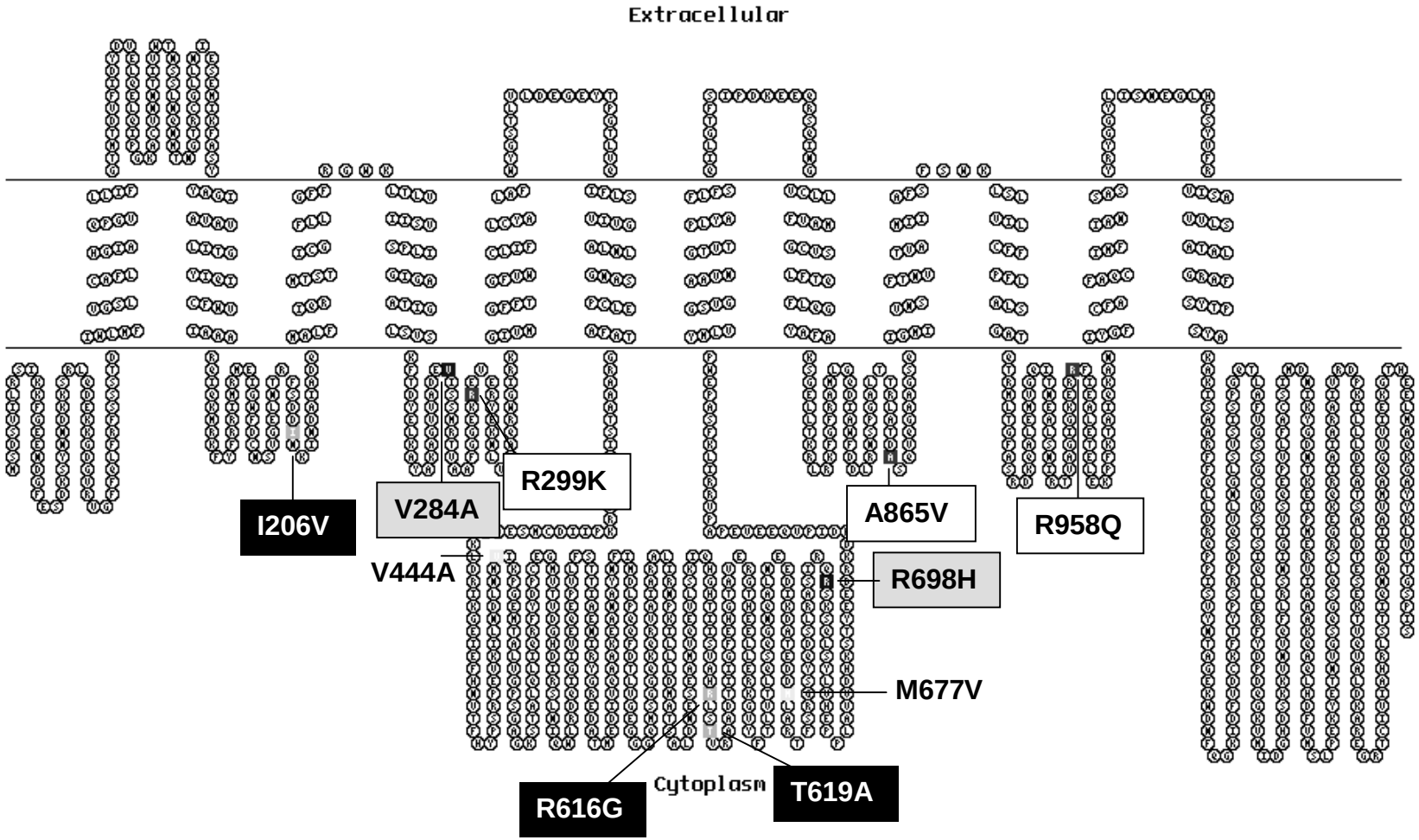

Fig. 2a

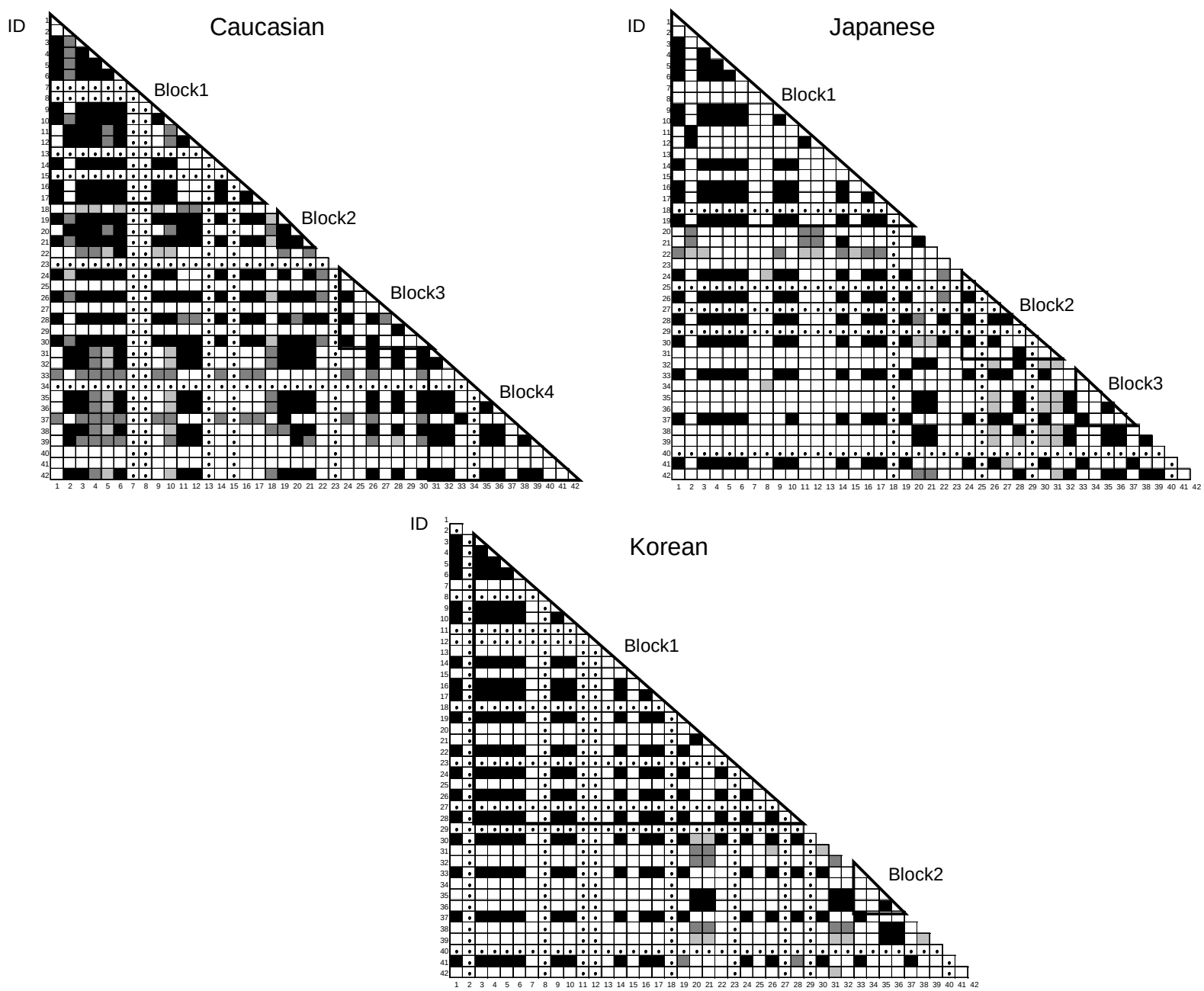

Fig. 2b

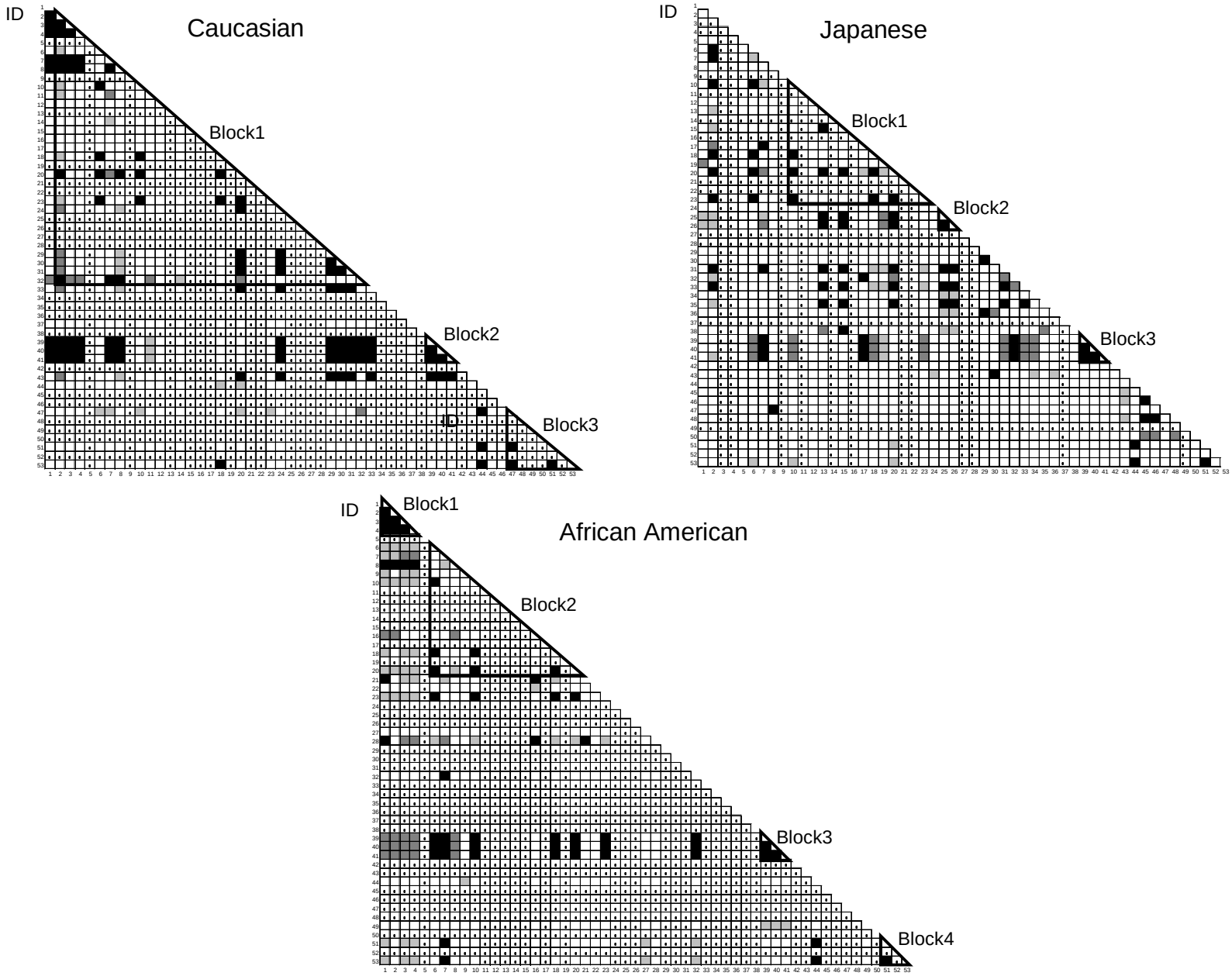

Fig. 3a

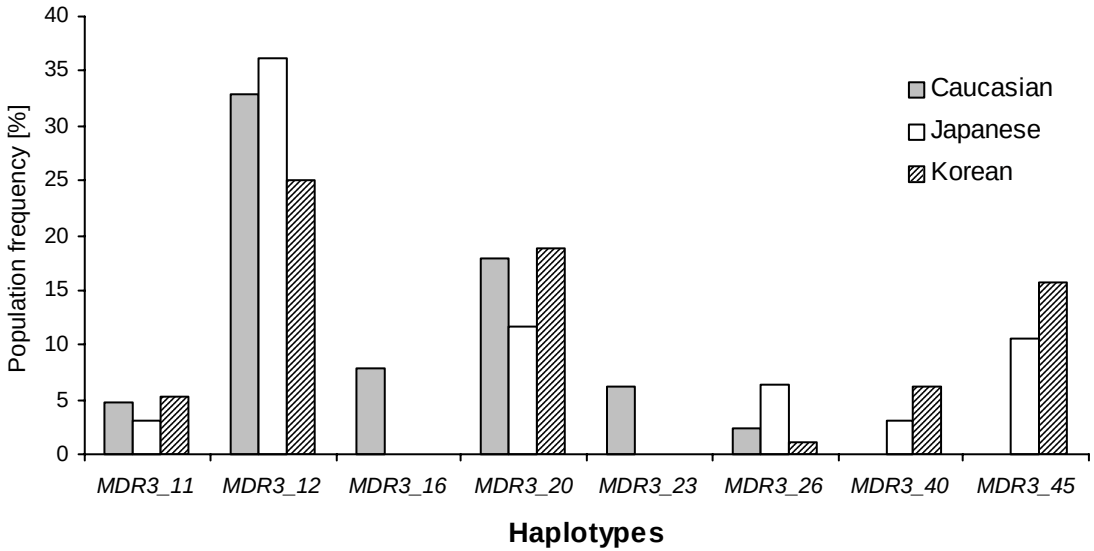

Fig. 3b

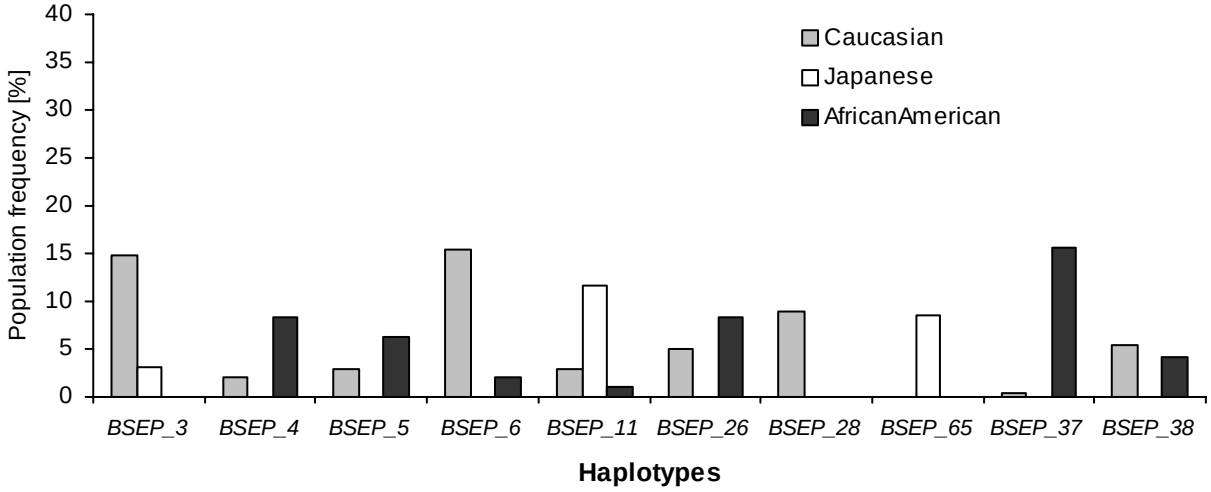

Fig. 4

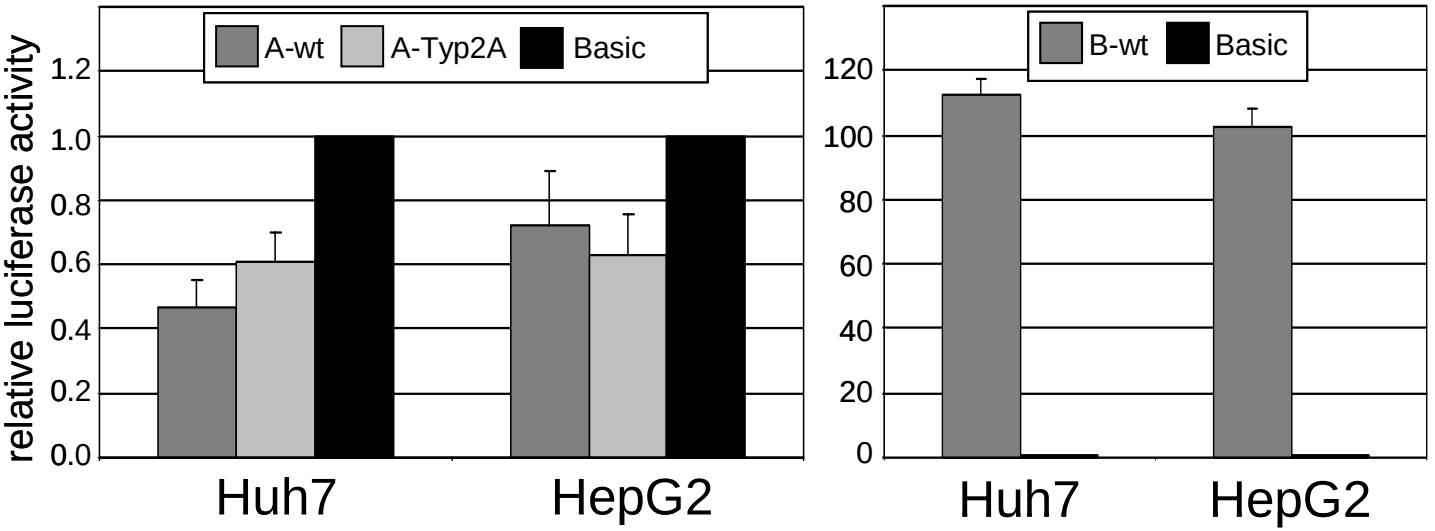

Fig. 5

Name	Variant ID	Position 1	-15281_-15278 delCTCT 2	-15150T>C 3	-15018G>A 4	-14944G>A 5	-14589A>T 55	-14524A>T 6	-14473G>A 7	-14437G>A 8	-14353T>C 9	-14312T>C 57	-14300A>T 10	-14207T>C 12	-14035A>G 58	-13910G>A 13	-13814G>A 15	-13662T>C 59	-13595G>A 83-86	-13523[T]n 17	-13297G>A
wt		wt	T	G	G	A	A	G	G	T	T	A	T	A	G	G	T	G	[T]12	G	
C_2		.	.	.	.	.	.	.	.	C	.	.	.	.	.	A	C	.	[T]9	.	
C_3		.	C	.	.	.	.	.	.	C	.	.	.	.	.	A	C	.	[T]9	.	
C_4		del	C	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	[T]11	.	
C_5		.	C	.	.	.	T	A	.	C	.	.	C	.	.	.	.	A	[T]10	.	
C_6		.	C	.	.	T	.	.	A	C	.	.	.	.	.	.	.	.	[T]11	.	
C_7		.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	[T]11	.	
C_8		.	C	.	.	.	.	.	A	C	.	.	.	.	A	.	.	.	[T]11	.	
C_9		.	C	.	.	.	.	A	.	C	.	T	.	.	.	.	.	.	[T]10	.	
C_10		.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	.	.	[T]10	.	
C_11		.	C	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	[T]10	.	
C_12		del	C	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	[T]11	A	
C_13		del	C	A	A	.	.	.	.	C	C	.	.	.	.	.	.	.	[T]11	.	
C_14		.	C	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	[T]11	.	

Fig. 6